Geophys. J. Int. (1990) 102, 25–43

Stochastic analysis of global traveltime data: mantle heterogeneity and random errors in the ISC data

O. Gudmundsson,* J. H. Davies† and R. W. Clayton
Seismological Laboratory, 252–21 California Institute of Technology, Pasadena, CA 91125, USA

Accepted 1989 December 18. Received 1989 December 12; in original form 1989 March 16

SUMMARY

Analysis of global traveltime data has been formulated in terms of the stochastic properties of the Earth's heterogeneity pattern and random errors in the data. The formalism relates the coherency of traveltime residuals within bundles of rays (summary rays) of varying size to the spherical harmonic power spectrum of the slowness field of the medium. It has been applied to mantle *P*-wave data from the ISC catalogue. The measure of coherency is the variance within summary rays. It is estimated within bins in source depth, epicentral distance and the scale size of the area defining a summary ray. The variance at infinitesimal scale length represents the incoherent component of the data (random errors). The variation of the variance with scale length contains information about the autocorrelation function or power spectrum of slowness perturbations within the Earth. The variation with epicentral distance reflects the depth variation of the spectrum. The formalism accounts for the uneven distribution (clustering) of stations and events.

We find that estimates of random errors correlate well with complexities on the traveltime curve of *P*-waves. The variance peaks at 1.0–2.0 s^2 at $\Delta \approx 20°$, where triplications occur on the traveltime curve, drops to 0.15–0.8 s^2 at teleseismic distances, and rises to 0.4–1.3 s^2 approaching the core shadow, where the traveltime curves of *P*-waves and *PcP*-waves merge. These estimates should be considered upper bounds for the random error variance of the data. The signal to random noise ratio in the teleseismic ISC *P*-wave data is about $S/N \approx 2$.

Inversion of the scale-dependent structural signal in the data yields models that concentrate heterogeneity strongly in the upper mantle. The product of correlation length and power drops by about two orders of magnitude from the surface of the Earth to the lower mantle. About half of this quantity in the upper mantle is due to small-scale features (<300 km). The lower mantle is devoid of small-scale structure. It contains 0.1 per cent velocity variations at a characteristic scale of about 1000 km. This corresponds to a spectral band-width of $l \approx 7$. The *D″* layer at the bottom 100–200 km of the mantle shows up as a distinct layer in our results. It has 0.3 per cent velocity variations at a characteristic scale of 350 km. The top of the lower mantle contains 0.3 per cent velocity variations on a scale of 500 km and also contains some small-scale power.

These results are compatible with previous deterministic lower mantle studies, although some details differ. The strength of heterogeneity in the upper mantle may obscure attempts to model the Earth's deep interior.

Key words: heterogeneity, mantle, power spectrum, random error, traveltime, variance.

* Now at: University of Iceland Science Institute, Dunhaga 5, 107 Reykjavik, Iceland.
† Now at: Department of Earth Sciences, University of Cambridge, Downing St., Cambridge CB2 3EQ, U.K.

INTRODUCTION

The International Seismological Centre (ISC) Catalogue contains over 20 yr of world-wide traveltime readings of

many seismic phases. This database of more than nine million picks is being actively used to map the aspherical structure of the Earth's interior on a global scale as well as regional lateral velocity variations near the Earth's surface. As time has progressed the global studies have sought to image ever deeper structure. Clayton & Comer (1983) and Dziewonski (1984) used 15 yr of ISC mantle *P*-wave data to map the lower mantle. Creager & Jordan (1986) and Morelli & Dziewonski (1987) used compressional core phases to map the core–mantle boundary. Morelli, Dziewonski & Woodhouse (1986) and Shearer, Toy & Orcutt (1988) used *PKIKP*-waves to map anisotropy in the inner core. The regional studies have concentrated on the structure of subducting slabs (e.g. Creager & Jordan 1985; Zhou & Clayton 1990). These studies together with surface wave tomography and long-period body wave synthesis provide important constraints on the workings of the geodynamo, and on the evolution of the Earth.

There are, however, some concerns about the ISC data and the techniques used to analyse them. The uncertainty of measurement is high compared with the signal attributable to aspherical structure. The data are contaminated by systematic errors due to misidentification of phases, earthquake mislocation, earthquake time function complexity, and potentially biased picking. The geometrical coverage of the data is uneven, due to the clustering of seismic sources in tectonically active regions and the lack of recording stations in the oceans. The images suffer from complex artifacts, which depend on the particular technique employed, due to uneven coverage. The strong small-scale velocity variations in the Earth's outermost layers are simplistically accounted for by station corrections. The severity of these problems is presently poorly understood. This has led much of the seismological community to view the results of global body wave tomography studies with scepticism.

In order to assess the quality of images obtained from the ISC data it is important to have robust estimates of the level of random errors in the data. Random errors can be propagated through inversion schemes to estimate their effect (Davies & Clayton 1987). This, however, gives only a lower bound on the model uncertainty, as systematic errors and model parametrization errors are ignored. It is not clear to what extent model amplitude exceeds random model error in some of the previous studies. (This will vary within each model).

Systematic errors are perhaps the most serious limitation of the ISC data. The effect of phase mispicks can be dealt with by windowing out cross overs in the traveltime curves. Gross errors are typically handled by excluding residuals exceeding a given value (4.0–5.0 s for compressional phases), thought to represent the maximum feasible structural signal. Beyond that, we are forced to assume that single station, single event or regionally systematic errors get randomly mixed within the model. This assumption is better justified in a stochastic spherically symmetric model than a deterministic laterally varying model.

Due to the size of global multidimensional inverse problems and the level of redundancy needed to suppress random scatter in the ISC data, models have been parametrized as truncated expansions in terms of continuous orthonormal functions [Dziewonski (1984), Morelli & Dziewonski (1987), and Morelli *et al.* (1986) used spherical harmonics and Legendre polynomials in depth] or in terms of discrete cells of fixed size (Clayton & Comer 1983; Humphreys, Clayton & Hager 1984; Zhou & Clayton 1990). This could give rise to significant truncation effects or aliasing of the small-scale signal, and is particularly worrisome since the variance reduction of previous studies has invariably been very low (5–15 per cent). It is not clear if summary rays effectively filter out signal contributions from scales smaller than their lateral extent because of the clustered geometry of both seismic sources and stations. It is not clear either that the remaining small-scale signal gets randomly mapped into the model. To understand this effect we need to know the power spectrum of the structure beyond the truncation wavenumber or harmonic degree. Creager & Jordan (1986) applied summary rays on a large scale ($20° \times 20°$) and thus obtained much higher variance reduction than other studies. They employed an intrinsic Gaussian spatial filter in their stochastic inverse scheme to suppress the aliasing problem.

It is well known from a variety of studies, ranging from array scattering studies to surface wave tomography studies, that the Earth's crust and upper mantle contain lateral velocity variations in excess of 5 per cent on length scales ranging from tens to thousands of kilometres. This exceeds the level of heterogeneity in models of the Earth's deep interior by an order of magnitude. Because of lack of resolution, teleseismic traveltime studies account for this region by applying static station (and event) corrections or by using differential traveltimes (Creager & Jordan 1987). Since small-scale velocity variations may extend to considerable depths (e.g. Aki 1973; Flatte & Wu 1988), where the separation of teleseismic rays to the same station may exceed the correlation length of the medium, a shallow small-scale signal is potentially mapped into models of the Earth's deep interior. Again a stochastic model describing the characteristic power spectrum of velocity variations provides sufficient information to study this effect synthetically.

We address some of these problems in this paper through a statistical approach to the interpretation of the data. We have formulated the analysis of global traveltime data in terms of the stochastic properties of the Earth's heterogeneity pattern and random errors in the data. The formalism relates the coherency of traveltime residuals within bundles of rays (summary rays) of varying size to the spherical harmonic power spectrum of the slowness field of the medium. The measure of coherency is the variance within summary rays. It is estimated within bins in source depth, epicentral distance and the scale size of the area defining a summary ray. The variation of the variance with scale length contains information about the autocorrelation function or power spectrum of slowness perturbations within the Earth. The variation with epicentral distance reflects the depth variation of the spectrum. The variance at infinitesimal scale length represents the incoherent component of the data (random errors). This stochastic approach allows for the separation of random variance (incoherent) from signal variance (coherent). It includes the entire volume through which the data pass and it seeks to explain all the variance in the data. While one can argue that spatially systematic errors in the ISC data are less likely

to have a significant effect in this stochastic approach than in deterministic inversions, because of the global averaging applied, the interpretation of incoherent variance as strictly random and coherent variance as due to structure is an assumption. We do not address the problem of systematic errors in this study.

In our view a stochastic model of the Earth's aspherical structure may yield more robust conclusions than previous deterministic models, but less specific. The loss of specificity inherent in the stochastic approach does however not lessen the usefulness of its conclusions as the modelling of mantle dynamics is still largely limited to idealized style or scale simulations.

DATA PROCESSING

In order to estimate summary ray variance as a function of epicentral distance, Δ, source depth, Z, and scale length Θ, we construct equal area grids covering the Earth's surface. The grids range in density from one cell covering the entire surface of the globe to one million cells. 18 different densities are used. The grids cells are rectangular in latitude–longitude coordinates (see Fig. 1). The grids possess a symmetry about the equator and the Greenwich longitudinal great circle. We use ellipticity corrected (Dziewonski & Gilbert 1976) ISC residuals from events with at least 50 picks reported. We omit events located at the surface or at a source depth of 33 km, because the ISC assigns poorly located events to these depths. Traveltime residuals larger than 4 s in absolute value are omitted. The data are binned in 2° windows of epicentral distance, ranging from 0° to 100° and six windows of source depth; 0–30 km, 30–60 km, 60–100 km, 100–200 km, 200–450 km and 450–650 km.

A summary ray is defined as the collection of all the data, which fall inside a given epicentral distance window and a given source depth window, and share both a source grid cell and receiver grid cell (see Fig. 1). The scale length is defined in terms of the scale angle, Θ

$$\Theta = \cos^{-1}\left(1 - \frac{2}{N}\right)$$

where N is the total number of cells in the grid. If we consider a circular area centred on the north pole of a sphere and measure its size by the colatitude of its margin, Θ, its area is given by $2\pi(1 - \cos\Theta)$. The grid cells in a grid with N cells are of area $4\pi/N$. Equating the two we get the

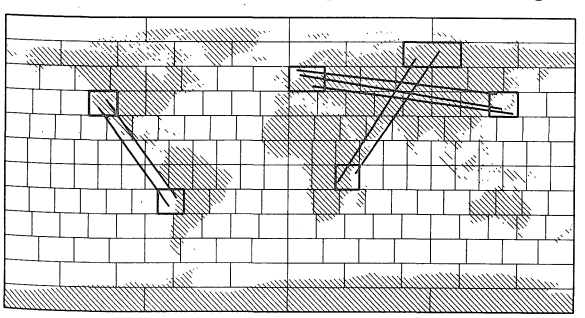


Figure 1. An example of the grids used to construct summary rays and three schematic sample summary rays. This grid has 192 cells or an angular measure of $\Theta = 8.3°$. The three sample summary rays are of similar length and could contribute to the same Δ–Z–Θ bin.

above definition of the scale angle. The variance within the kth summary ray is calculated as the variance of its residuals referenced to their mean.

$$\sigma_k^2 = \sum_{i=1}^{n_k} (\delta t_i - \overline{\delta t}_k)^2/(n_k - 1)$$

where n_k is the number of data in the kth summary ray, δt_i are the individual time residuals and $\overline{\delta t}_k$ is the mean residual for the summary ray. A summary ray is included only if $n_k \geq 4$. The variance estimate for each scale, distance, and depth combination, is taken to be the average of the variances of all the summary rays found for that combination. Finally, each grid is rotated four times in equal increments that span the width of the grid cells, and the variance estimates averaged. At the small-scale end of the spectrum, $\Theta < 5^0$, estimates of variance based on fewer than 25 summary rays are omitted.

A reference ray is ascribed to each Δ–Z bin. It is defined by the average epicentral distance and average source depth of the composite rays. A ray parameter is calculated by tracing through the JB (Jeffreys & Bullen) model.

The variance estimates are plotted as functions of scale length for each pair of Δ and Z. Some examples are shown in Fig. 2. The behaviour of the data is systematic. The curves rise rapidly at small scales and become relatively flat at large scales. The flat level of the curves invariably decreases with source depth. The behaviour with epicentral distance is more complex. At teleseismic distances there is a gradual decrease with Δ out to about 90°, where the variance starts increasing again. Fig. 3 shows examples of the behaviour at small scales. Extrapolation of the curves to vanishing scales does not yield a zero intercept with the variance (vertical) axis.

The overall behaviour of the data is in accordance with the formalism developed in the following section. A non-zero intercept is expected due to picking errors, event mislocations, the finite binning of the data (particularly in source depth), and the lack of scale resolution at small scales. A finite correlation length characterization of the medium predicts the general shape of the curves (see Fig. 2). As the cross-section of the summary ray tube increases, more and more component rays travel through uncorrelated slowness anomalies and the variance increases. As source depth increases the length of the path within the heterogeneous Earth decreases and the variance decreases. If the Earth were uniformly heterogeneous the variance would increase as the total path length increased and hence as epicentral distance increased. The overall decrease in variance with epicentral distance at teleseismic distances thus implies that the strength of heterogeneity is concentrated at shallow depths (see Fig. 4). The small, but consistent, drop in variance at $\Theta = 60°$–$90°$ at shallow source depths does not fit in with the above scenario. We cannot explain this feature, but suggest that it is due to biases in sampling. While at small scales the sampling is dominated by continental and tectonic regions, oceanic regions come increasingly into play at large scales, but only at shallow source depths. If the small-scale structure of oceanic crust and upper mantle is weaker than that of continental and tectonic regions, the above effect would be produced.

The intercepts of the individual curves in Figs 2 and 3 with the variance axis represent incoherent variance or random

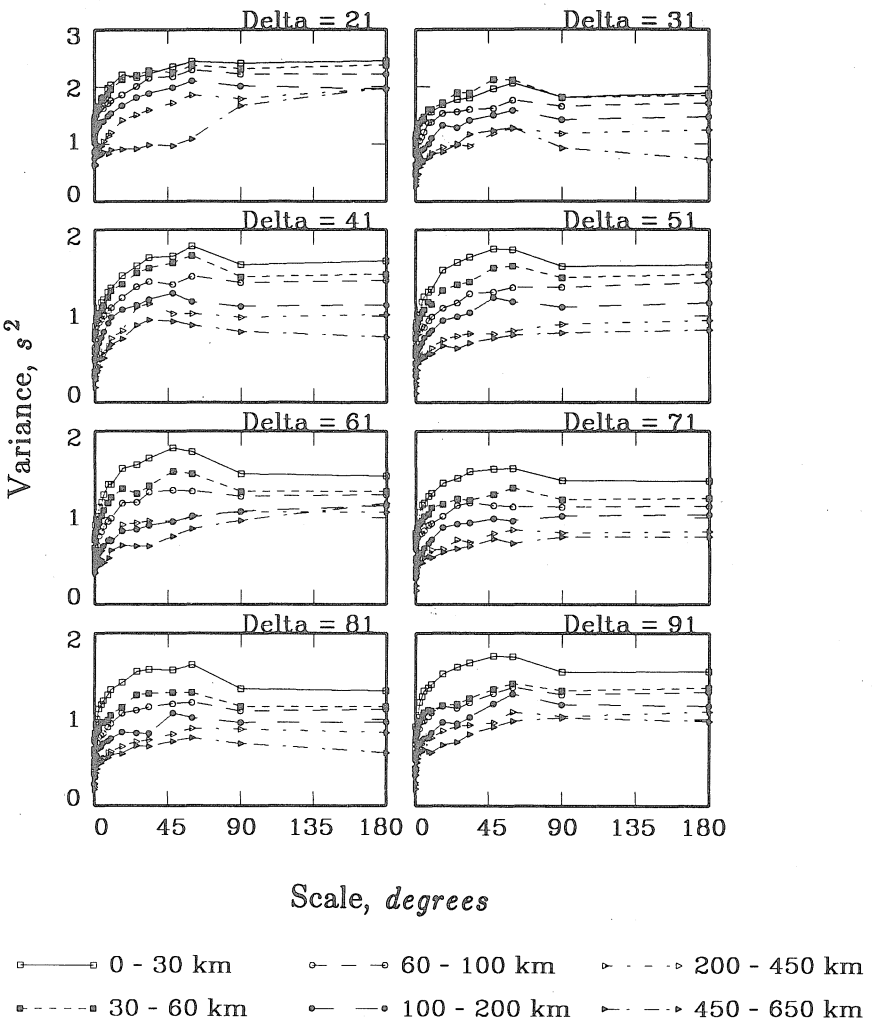


Figure 2. Samples of data as functions of scale in the distance range $\Delta = 20°–90°$ as indicated. The six curves in each frame represent the six depth bins. The shallowest depth bin consistently has the highest variance and the deepest one the lowest variance.

errors independent of the structure of the Earth. We estimate the intercepts by a constrained linear extrapolation to the variance axis. We calculate a least-squares linear fit through the small-scale data, $\Theta < 1.5°$, but constrain the slope to be inversely proportional to $\cos(i)$, where i is the angle of incidence, for all curves at a common source depth, i.e. the curves that share a source depth bin have a slope, b, described by the common constant, B, according to $b = B/\cos(i)$. The behaviour of the data at small scales is generally linear as is evident from Fig. 3. We assume that the small-scale signal (the slope at the origin), is primarily accumulated at shallow depths, where the path length is inversely proportional to the cosine of the angle of incidence. Hence the form of the constraint. The constraint is needed in order to minimize the effects of the potential biases and errors at small scales. The resulting intercept estimates are plotted as a function of epicentral distance in Fig. 5. The behaviour is similar for all six source depth bins. The static variance peaks sharply in the triplication region ($15° < \Delta < 25°$), stays relatively flat at teleseismic distances, and increases beyond $\Delta = 90°$ as the traveltime curves for *P*- and *PcP*-waves merge. The static variance decreases with source depth. The behaviour of the static variance with epicentral distance may be attributed to picking errors since it is high where complexities occur on the traveltime curve. The behaviour with source depth may be partially attributed to the variation of picking errors or location errors with depth, but is likely to have a contribution from the finiteness of the source depth binning and the lack of scale resolution, due to the very fine structure of the earth (scale lengths of the order of tens of km).

We use the smoothed description of the intercepts as our estimates of random error (see Fig. 5b). We then remove the random variance from the data. The remaining variance may be interpreted in terms of statistical measures of the earth's slowness field as formulated in the following section. Fig. 4 summarizes the behaviour of the reduced data with epicentral distance. Plotted is the asymptotic variance level at large scales.

THEORY

The application of stochastic structural modelling in seismology has been limited. It has been restricted to the analysis of dense array data, notably from LASA and NORSAR. A number of workers have applied the theory

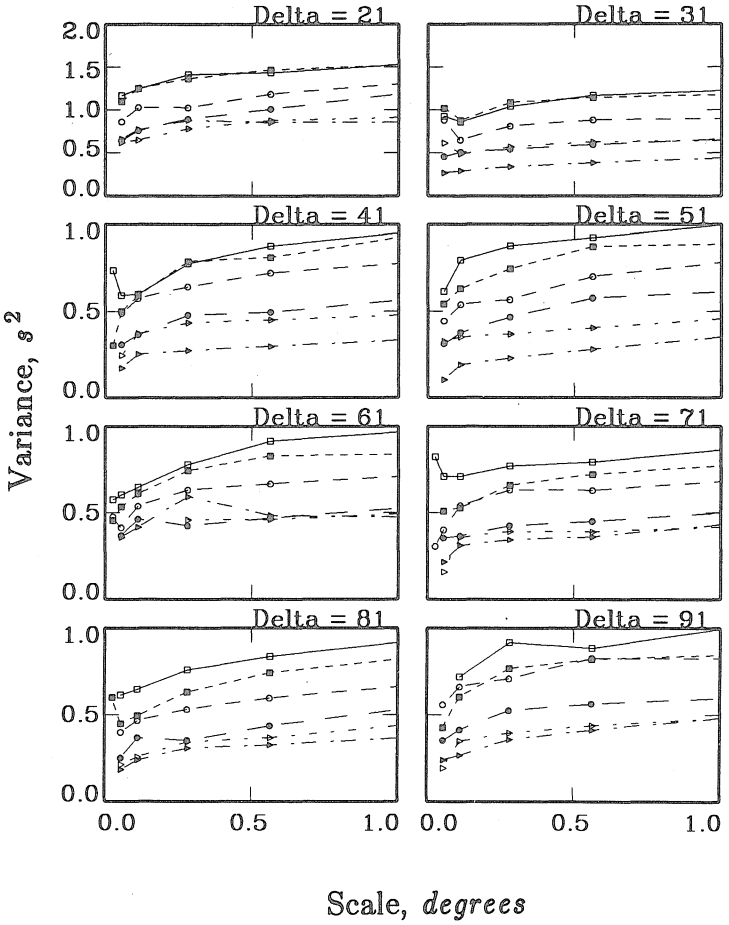


Figure 3. Same as Fig. 2 except showing only the data at very small scales. Symbol legend same as in Fig. 2.

developed by Chernov (1960) to study the heterogeneous character of the lithosphere and upper mantle (e.g. Aki 1973; Capon 1974; Berteussen *et al.* 1975a,b). Flatte & Wu (1988) used a somewhat simpler approach based on the parabolic wave equation, which allows for more flexibility in the modelling. Frankel & Clayton (1986) used a synthetic approach to this problem, utilizing the finite difference method. Other problems that have been analysed with some success by stochastic means include the generation of coda by crustal heterogeneity (e.g. Aki & Chouet 1975; Gao *et*

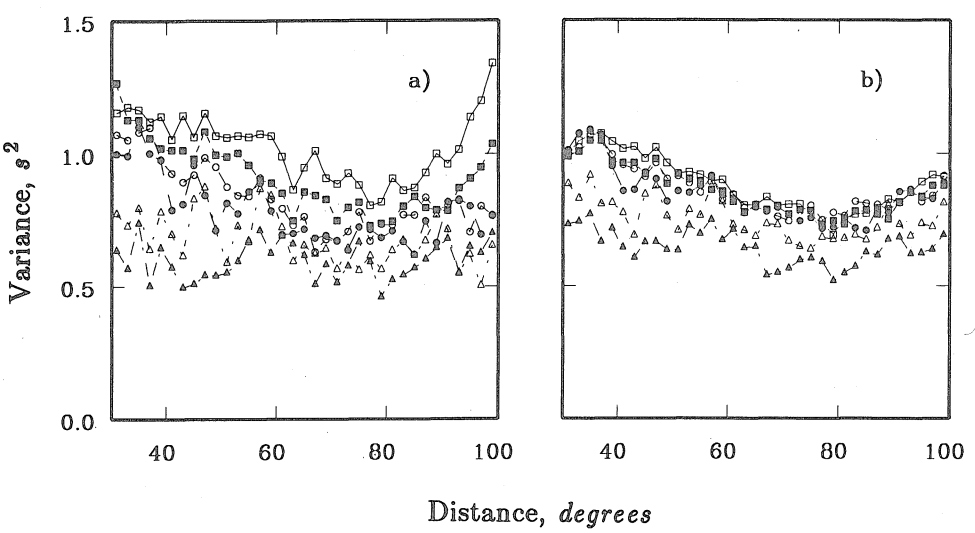


Figure 4. The asymptotic level of variance at large scales as a function of epicentral distance for the six depth bins. (a) Using smoothed intercept estimates (Fig. 5b); (b) using intercepts in Fig. 5(c). Note the reduction in variance with source depth, the negative slope out to 80°, and the positive slope beyond 85°. Symbol legend same as in Fig. 2.

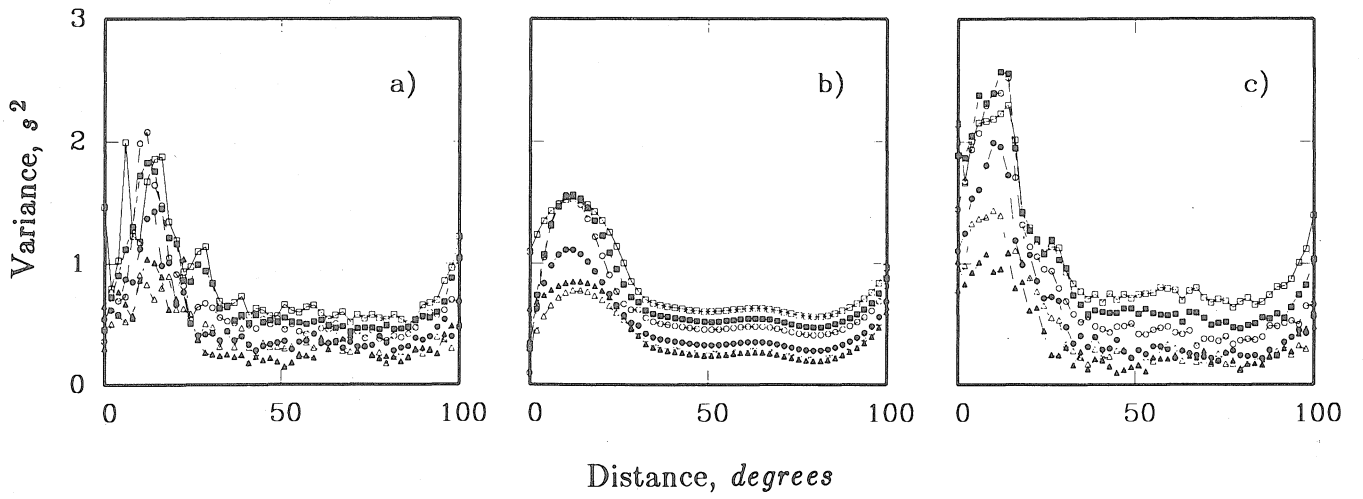


Figure 5. Estimates of intercepts or incoherent error variance as a function of epicentral distance. (a) Independent estimates; (b) smoothed by polynomial fitting; (c) including systematic misfit of model to data (see Inversion and Results). The six curves represent the six source depth bins as before. Symbol legend same as in Fig. 2.

al. 1983a,b) and the generation of *PKP* precursors by heterogeneity near the core–mantle boundary (e.g. Doornbos & Vlaar 1973; Haddon & Cleary 1974).

We take a simpler approach to a greater volume of poorer data than the above studies, but apply many of the same concepts, i.e. those of a random medium. A full wave theoretical (scattering) formulation is inapplicable since we are not dealing with single wavefronts. Yet, the diffraction effects described by scattering theory occur since the data are traveltime readings of finite frequency seismograms. The most important such effect is probably wavefront healing, which may be systematic in making the medium appear faster than its intrinsic properties imply (Wielandt 1987).
We take a somewhat simplistic, but pragmatic, attitude in regarding the finite wavelength property as an inherent limitation to the data, rendering them intrinsically insensitive to the small-scale features of the medium. The resulting characterization of the medium may then be interpreted as an apparent property, reflecting the true property intrinsically smoothed. This implies the existence of an inner scale of sensitivity of the data, which for short-period teleseismic waves would be of the order of 10 km.

Our approach is a ray theoretical approach, which may be regarded as an adaptation of the linearized inverse problem of traveltime residuals to the concepts of stochastic analysis. It is designed to retrieve information about the Earth's mantle slowness field from traveltime data as compiled in the previous section. First we develop a formalism for a Cartesian infinite world in some detail and then adapt that to the finite spherical Earth.

Consider the Earth to contain a random function, $\delta U(\mathbf{x})$, of small slowness perturbations. Let δU be characterized by an autocorrelation function, $R(\mathbf{x}, \mathbf{y})$. Here $\mathbf{x}$ is a position vector within the earth and $\mathbf{y}$ is the displacement vector between two arbitrary points. Assume that the statistics of $\delta U(\mathbf{x})$ are isotropic and spherically symmetric, such that $R(\mathbf{x}, \mathbf{y}) = R(r, y)$, where r is distance from the Earth's centre and y is the length of the vector $\mathbf{y}$. Next consider a projection of the medium on to the earth's surface by means of curvilinear integrals to be specified later, i.e. that we have a random function $\delta t(\xi)$, where ξ is the position vector on the surface. Let this function be characterized by the autocorrelation function $T(\xi, \eta)$, where η is a displacement vector. Since δU is isotropic and spherically symmetric, we can assume that δt is isotropic and stationary, if the mapping of δU to δt is one to one or uniformly random. Thus $T(\xi, \eta) = T(\eta)$, where η is the length of the displacement vector η. We need to estimate the scatter within a sample of circular area $A = \pi a^2$ of δt, i.e. the variance of all samples within A referenced to the mean of the area.

Assume that the global mean of δt is zero. Then the variance is

$$\begin{aligned}\sigma^2(a) &= E\{[\delta t(\xi) - \overline{\delta t}(a)]^2\} \\ &= E\{\delta t^2(\xi)\} - E\{\overline{\delta t}^2(a)\}.\end{aligned} \tag{1}$$

Here $E\{\}$ stands for the statistical expectation over A and $\delta t(a)$ for the mean of δt within A. The first term in equation (1) is simply the global variance of δt, i.e.

$$E\{\delta t^2(\xi)\} = T(0). \tag{2}$$

The second term in equation (1) is the variance of the mean

$$\begin{aligned}E\{\overline{\delta t}^2(a)\} &= E\left\{\left[\frac{1}{\pi a^2}\int_A b_a(\xi)\,\delta t(\xi)\,d\xi\right]^2\right\} \\ &= \frac{1}{\pi^2 a^4}\int_A\int E\{\delta t(\xi_1)\,\delta t(\xi_2)\} \\ &\quad \times E\{b_a(\xi_1) b_a(\xi_2)\}\,d\xi_1\,d\xi_2 \\ &= \frac{1}{\pi^2 a^4}\int_A\int T(|\xi_1 - \xi_2|) B(a, |\xi_1 - \xi_2|)\,d\xi_1\,d\xi_2\end{aligned} \tag{3}$$

where $b_a(\xi)$ is a scaled sampling function (an irregular 2-D comb function).

$$b_a(\xi) = \frac{\pi a^2}{N}\sum_{i=1}^{N}\delta(\xi_i) = \frac{1}{\eta_a}\sum_{i=1}^{N}\delta(\xi_i)$$

where N is the number of measurements within the region A, η_a is the average sample density within A, and ξ_i are the

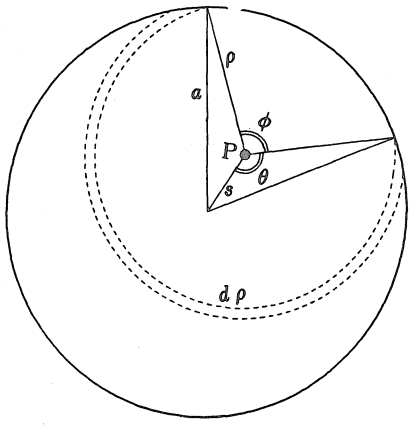


Figure 6. Geometrical definitions of the parameters used in the derivation of equation (6).

sample positions. We can equivalently think of b_a as a normalized random sample density. The function $B(a, \rho)$ is then the autocorrelation of the sampling density. By separating the expectation of the sampling and the structure we have assumed that the two are independent. In other words, it is assumed that the source and receiver locations are not correlated with structural anomalies. This assumption is suspect for subduction events.

Consider a point $\xi_1 = P$ a distance s from the centre of A, and all points ξ_2 a distance ρ from P, within A (see Fig. 6). If $\rho < (a - s)$ the points ξ_2 span a whole circle. If $(a - s) < p < (a + s)$ the angle ϕ is excluded. The cosine rule gives

$$a^2 = s^2 + \rho^2 - 2s\rho \cos(\theta). \tag{4}$$

Thus the arc spanned by the points ξ_2 is of length $2\theta = 2\pi$ if $0 < \rho < a$ and $0 < s < (a - \rho)$, and

$$2\theta = 2\cos^{-1}\left(\frac{s^2 + \rho^2 - a^2}{2\rho s}\right)$$

if $0 < \rho < a$ and $(a - \rho) < s < a$ or if $a < \rho < 2a$ and $(\rho - a) < s < a$. Since δt is isotropic, T is a function of ρ only, and assuming that the sampling density is isotropic, we can rewrite (3) as

$$\begin{aligned} E\{\overline{\delta t}^2(a)\} = \frac{1}{\pi^2 a^4}\Bigg[& \int_0^a 2\pi\rho T(\rho)B(a, \rho)\, d\rho \int_0^{a-\rho} 2\pi s\, ds \\ & + \int_0^a 2\pi\rho T(\rho)B(a, \rho)\, d\rho \\ & \times \int_{a-\rho}^{a} 2\cos^{-1}\left(\frac{s^2 + \rho^2 - a^2}{2\rho s}\right) s\, ds \\ & + \int_a^{2a} 2\pi\rho T(\rho)B(a, \rho)\, d\rho \\ & \times \int_{\rho-a}^{a} 2\cos^{-1}\left(\frac{s^2 + \rho^2 - a^2}{2\rho s}\right) s\, ds \Bigg] \end{aligned} \tag{5}$$

which after simple, but tedious algebra simplifies to

$$E\{\overline{\delta t}^2(a)\} = \int_0^{2a} \omega(a, \rho)T(\rho)B(a, \rho)\, d\rho \tag{6}$$

where

$$\omega(a, \rho) = \frac{4}{\pi a^2}\left\{\cos^{-1}\left(\frac{\rho}{2a}\right) - \frac{\rho}{2a}\left[1 - \left(\frac{\rho}{2a}\right)^2\right]^{1/2}\right\}\rho.$$

Equation (6) describes the variance of the average $\overline{\delta t}$ over area of radius a. It is informative to consider a few simple cases. Assume that the sampling is uniform. Then $B = 1$. If the autocorrelation is a constant within the area, $T(\rho) = T_0$, the variance of the mean is T_0. This is expected since in this case δt is constant within the area and comparing randomly selected areal averages is equivalent to comparing randomly selected points on δt. The variance of the averages should thus be the variance of the function itself, $T(0) = T_0$. From the above it is clear that if $T(\rho)$ has a zero derivative at the origin, we get

$$E\{\overline{\delta t}^2(0)\} = T(0). \tag{7}$$

An other interesting example is a completely uncorrelated δt function, i.e. $T(\rho) = \delta(\rho)/\rho$, where δ represents the Dirac delta function. In this case the variance of the averages decays as the inverse of the area (analogous to the $1/N$ decay in the case of discrete sampling, where N is the number of samples).

Figure 7 shows some examples of the weighting functions of equation (6), $\omega(a, \rho)$. Their integral is unity and they go to zero at the endpoints $(0, 2a)$. They peak just below the centre of their domain. Thus, when averaging uniformly over a circle of radius a, the main contribution to the variance is from a scale length of $\rho = a$.

Substituting equations (2) and (6) into equation (1) we get

$$\begin{aligned} \sigma^2(a) &= T(0) - \int_0^{2a} \omega(a, \rho)T(\rho)B(a, \rho)\, d\rho \\ &= \int_0^{2a} \omega(a, \rho)B(a, \rho)[T(0) - T(\rho)]\, d\rho \end{aligned} \tag{8}$$

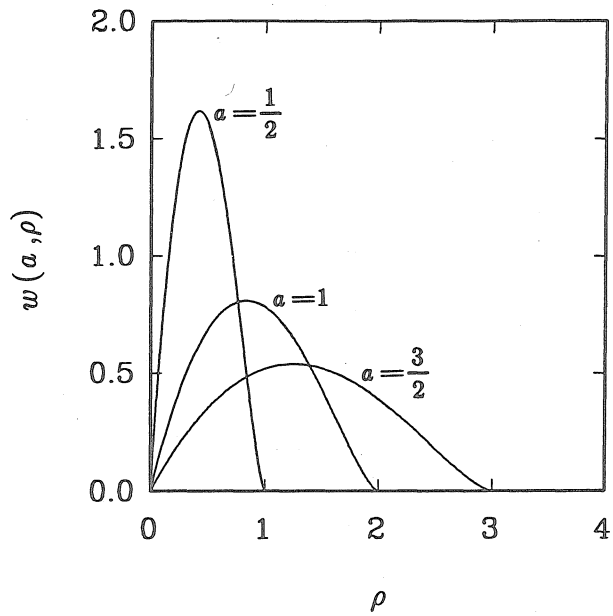


Figure 7. The weighting functions in equation (6), $\omega(a, p)$, for a few choices of scale length.

since

$$\int_0^{2a} \omega(a, \rho) B(a, \rho)\, d\rho = 1$$

as ωB is the probability density of sample spacing within a circular region of radius a. Equation (8) describes the variance within area $A = \pi a^2$, referenced to the mean within A.

At this point we have a tool to relate our observations of the variance within summary rays to the autocorrelation function of the traveltime function. We need to express that in terms of the autocorrelation of the medium. In order to do that we must idealize the summary ray geometry, i.e. assume a specific deterministic or random geometry for the rays contributing to a summary ray.

We calculate the variance of summary rays subject to fine binning in both source depth and epicentral distance. Thus, even when the averaging area becomes large, the ray parameter and hence the ray geometry stays relatively constant within each summary ray. Consequently, the contributing summary rays are not dominated by fans of rays stemming from a single receiver or a single event, except at the smallest scale lengths. It is thus reasonable to assume that all the rays have the same ray parameter and randomly distributed endpoints in the two grid cells defining the summary ray. This implies that the rays are approximately parallel and simply shifted horizontally relative to each other. We can write

$$\delta t(\xi) = \int_{\text{ray}} \delta U(\mathbb{x})\, dS. \tag{9}$$

Thus

$$\begin{aligned} T(\rho) &= E\{\delta t(\xi_1)\, \delta t(\xi_2)\} \\ &= \int_{\text{ray}_1} \int_{\text{ray}_2} E\{\delta U(\mathbb{x}_1)\, \delta U(\mathbb{x}_2)\}\, dS_1\, dS_2 \\ &= \int_{\text{ray}_1} \int_{\text{ray}_2} R(|\mathbb{x}_1 - \mathbb{x}_2|)\, dS_1\, dS_2. \end{aligned} \tag{10}$$

If we assume that the statistics of δU are a weak function of position within the earth and that the correlation length is small compared to the radius of curvature of the rays, we can think of the rays as locally straight and the autocorrelation function as locally constant. If ρ is the horizontal distance between two rays, i the angle of incidence and ϕ the angle the ray displacement vector makes with the ray azimuth, then the minimum distance between the two rays is (see Fig. 8)

$$d = \rho \sqrt{\cos^2(\phi) \cos^2(i) + \sin^2(\phi)}. \tag{11}$$

If we fix our attention to the point P_1 on ray 1 at Z and define S as ray length along ray 2, passing through 0 at P_2, the point on ray 2 closest to P_1, i.e. distance d away, we can write

$$\tau = \sqrt{d^2 + S^2}. \tag{12}$$

Hence

$$|S| = \sqrt{\tau^2 - d^2} \tag{13}$$

and

$$|dS| = |dS_2| = \frac{\tau\, d\tau}{\sqrt{\tau^2 - d^2}}. \tag{14}$$

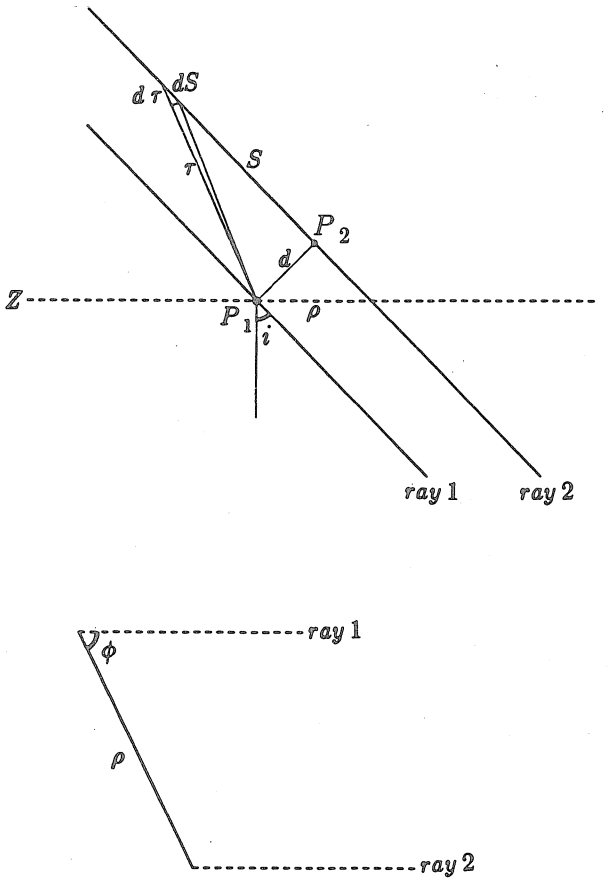


Figure 8. Geometrical definitions of the parameters used in the derivation of equation (17). The top panel is a vertical cross-section, the lower one a map view.

The integral over dS_2 (ray 2) in equation (10) can then be translated into an integral over τ, taken from d to ∞.

$$T(\rho) = 2 \int_{\text{ray}_1} \int_d^\infty \frac{\tau R(\tau)\, d\tau}{\sqrt{\tau^2 - d^2}}\, dS_1. \tag{15}$$

According to equation (11) the close point, d, depends on the azimuth of the ray displacement vector. For nearly vertical rays, $\theta = 0$, we have $d \approx \rho$. At the turning point the deviation of d from ρ is at its maximum with $d = |\sin \phi|\, \rho$, which on the average is $d = 2\rho/\pi$. The variation of d is thus not a strong effect, particularly if the correlation length is large or the level of heterogeneity is small at the turning point and we will simply approximate d by ρ.

If we now assume that the autocorrelation function of the medium is Gaussian, $R(\tau) = R_0 e^{-\tau^2/\alpha^2}$, we can simplify equation (15):

$$\begin{aligned} T(\rho) &= 2 \int_{\text{ray}} \int_\rho^\infty R_0 \frac{\tau e^{-\tau^2/\alpha^2}\, d\tau}{\sqrt{\tau^2 - \rho^2}}\, dS \\ &= \sqrt{\pi} \int_{\text{ray}} R_0 \alpha e^{-\rho^2/\alpha^2}\, dS. \end{aligned} \tag{16}$$

This result holds approximately for many choices of an autocorrelation function, provided it does not have strong side lobes, i.e. the structure of the medium must not have a strong periodic component. Stated in a slightly different

form, equation (16) becomes

$$T(\rho) \approx C \int_{\text{ray}} x_{1/2} R(\rho)\, dS \tag{17}$$

where $x_{1/2}$ is the half-width of $R(\tau)$, $[R(x_{1/2}) = 0.5R(0)]$, and C is a constant of order 2.0. For a Gaussian autocorrelation function, equation (17) holds exactly and the constant is

$$C = \sqrt{\frac{\pi}{\ln 2}} = 2.129.$$

If R_0 and α are constant along the ray path we get

$$T(\rho) = \sqrt{\pi} R_0 \alpha S e^{-\rho^2/\alpha^2} \tag{18}$$

which agrees with the transverse autocorrelation of phase of Chernov (1960) in the high-frequency limit. Chernov derived his results by a wave theoretical approach and showed that when the wave parameter, $D = 4S/k\alpha^2$, is not negligibly small ($D \ll 1$), diffraction effects (wavefront healing) cannot be ignored. (Here k stands for wavenumber, S for path length and α for correlation length.) Wavefront healing describes a physical smoothing process on a single wavefront. We do not isolate single wavefronts in this study and cannot account for this effect properly.

We can combine equations (8) and (17) to write the observables directly in terms of the parameters that we seek, namely $R(r, \tau)$

$$\begin{aligned}\sigma^2(a) &= C \int_{\text{ray}} x_{1/2}(r) \\ &\quad \times \left[R(r, 0) - \int_0^{2a} \omega(a, \rho) B(a, \rho) R(r, \rho)\, d\rho \right] dS \\ &= C \int_{\text{ray}} x_{1/2}(r) \\ &\quad \times \int_0^{2a} \omega(a, \rho) B(a, \rho) [R(r, 0) - R(r, \rho)]\, d\rho\, dS.\end{aligned} \tag{19}$$

The above formulation assumes the Earth to be Cartesian and flat. It is applicable to localized studies. For the present global study we need to account for the finiteness and sphericity of the Earth.

Let the medium be perturbed by

$$dU(r, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} A_{lm}(r) Y_{lm}(\theta, \phi) \tag{20}$$

where $Y_{lm}(\theta, \phi)$ are fully normalized spherical harmonics and θ is colatitude, ϕ is longitude and $A_{lm}(r)$ are harmonic coefficients varying with Earth radius.

$$Y_{lm}(\theta, \phi) = \left(\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!} \right)^{1/2} P_l^m(\cos\theta) e^{im\phi} \tag{21}$$

where P_l^m are the associated Legendre functions. We take the harmonic coefficients to be independent random functions of depth. We can then define an autocorrelation function in terms of the power spectral coefficients of the structure, Q_l. By definition

$$Q_l = \frac{1}{(2l+1)} \sum_{m=-l}^{l} A_{lm} A_{lm}^* = E\{A_{lm} A_{lm}^*\} \tag{22}$$

for a given set of surface spherical harmonics. Here * represents the complex conjugate. Allowing for some coherency in the harmonic pattern with depth we thus define

$$E\{A_{lm}(r_1) A_{pq}^*(r_2)\} = Q_l\left(\frac{r_1 + r_2}{2}\right) f(|r_1 - r_2|)\, \delta_{lp}\, \delta_{mq} \tag{23}$$

where f describes the depth correlation and δ is the Kronecker delta operator. Thus correlating the traveltime residuals for two rays we get

$$\begin{aligned}&E\{\delta t_1\, \sigma t_2\} \\ &= E\Bigg\{ \int_{\text{ray}_1} \int_{\text{ray}_2} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \sum_{p=0}^{\infty} \sum_{q=-p}^{p} A_{lm}(r_1) A_{pq}^*(r_2) \\ &\quad \times \left(\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!} \right)^{1/2} \left(\frac{2p+1}{4\pi} \frac{(p-q)!}{(p+q)!} \right)^{1/2} \\ &\quad \times P_l^m(\cos\theta_1) P_p^q(\cos\theta_2) e^{im\phi_1} (e^{iq\phi_2})^*\, dS_1\, dS_2 \Bigg\} \\ &= \int_{\text{ray}_1} \int_{\text{ray}_2} \sum_{i=0}^{\infty} \sum_{m=-l}^{l} Q_l\left(\frac{r_1 + r_2}{2}\right) f(|r_1 - r_2|) \frac{2l+1}{4\pi} \\ &\quad \times \frac{(l-m)!}{(l+m)!} P_l^m(\cos\theta_1) P_l^m(\cos\theta_2) e^{im(\phi_1 - \phi_2)}\, dS_1\, dS_2.\end{aligned} \tag{24}$$

The latter identity comes about by substituting equation (23) into equation (24). Using the addition theorem for associated Legendre functions we can simplify equation (24) to

$$\begin{aligned}&E\{\delta t_1\, \delta t_2\} \\ &= \frac{1}{4\pi} \int_{\text{ray}_1} \int_{\text{ray}_2} \sum_{l=0}^{\infty} (2l+1) Q_l(r) f(\delta r) P_l(\cos\lambda)\, dS_1\, dS_2\end{aligned} \tag{25}$$

where

$$\cos\lambda = \cos\theta_1 \cos\theta_2 \cos(\phi_1 - \phi_2) + \sin\theta_1 \sin\theta_2,$$

$$r = (r_1 + r_2)/2,$$

$$\delta r = r_1 - r_2.$$

λ is the angular distance between two arbitrary points on the two rays. Equation (25) is analogous to equation (10) for the Cartesian case, except it is cast in the spectral domain. The integrand

$$\frac{1}{4\pi} \sum_{i=0}^{\infty} (2l+1) Q_l(r) f(\delta r) P_l(\cos\lambda)$$

may be thought of as an autocorrelation function for the medium. To proceed we make the same assumptions as before; the two rays stay parallel along their entire path, the statistics of the medium are isotropic, the statistics vary smoothly along the path, the propagation distance far exceeds the correlation length and the radius of curvature of the rays is much greater than the correlation length. We can then apply the same arguments as in equations (13)–(18) to conclude

$$T(\lambda) = \frac{C}{4\pi} \int_{\text{ray}} x_{1/2}(r) \sum_{l=0}^{\infty} (2l+1) Q_l(r) P_l(\cos\lambda)\, dS \tag{26}$$

where T is the autocorrelation of the travetime residuals as mapped on to the earth's surface and λ is the horizontal angular distance between the two rays.

We proceed in the same manner as in equations (1)–(6) to

derive the effect of areal averaging on a spherical surface. Let Θ represent the angular radius of the area at hand (if the area is centred on the pole, Θ is the colatitude of its margin). Let λ represent angular distance between two arbitrary points within the area. Then the area is

$$A = 2\pi(1 - \cos\Theta)$$

and provided $\Theta < \pi/2$ the areal weighting function ω is

$$\omega(\Theta, \lambda) = \frac{\pi - 4\cos\Theta\cos^{-1}\alpha + \cos^{-1}\beta_1 + \cos^{-1}\beta_2}{2\pi(1-\cos\Theta)^2}\sin\lambda \qquad (27)$$

if $0 < \lambda < \Theta$ and

$$\omega(\Theta, \lambda) = \frac{\pi - 4\cos\Theta\cos^{-1}\alpha + \sin^{-1}\beta_1 - \sin^{-1}\beta_2}{2\pi(1-\cos\Theta)^2}\sin\lambda$$

if $\Theta < \lambda < 2\Theta$, where

$$\alpha = \frac{\cos\Theta(1-\cos\lambda)}{\sin\Theta\sin\lambda},$$

$$\beta_1 = \frac{(1-\cos\lambda)[1+\cos\lambda-\cos\Theta(1+\cos\Theta)]}{(1-\cos\Theta)\sin\Theta\sin\lambda}$$

and

$$\beta_2 = \frac{(1-\cos\lambda)[1+\cos\lambda+\cos\Theta(1-\cos\Theta)]}{(1+\cos\Theta)\sin\Theta\sin\lambda}.$$

We have not found a closed form solution for $\Theta > \pi/2$, except at $\Theta = \pi$ the solution is $[\omega(\pi, \lambda) = 1/2[\sin(\lambda)]$. In that case this areal averaging kernel must be evaluated numerically. Some examples are shown in Fig. 9.

We can now put together an expression relating the observable, $\sigma^2(\Theta)$ to the structural parameters that we seek for the spherical finite Earth by combining equations (8), (26) and (27).

$$\sigma^2(\Theta) = \frac{C}{4\pi}\sum_{l=0}^{\infty}\int_0^{2\Theta}\int_{\text{ray}}\omega(\Theta,\lambda)B(\Theta,\lambda) \times (2l+1)x_{1/2}(r)Q_l(r)[1-P_l(\cos\lambda)]\,dS\,d\lambda \qquad (28)$$

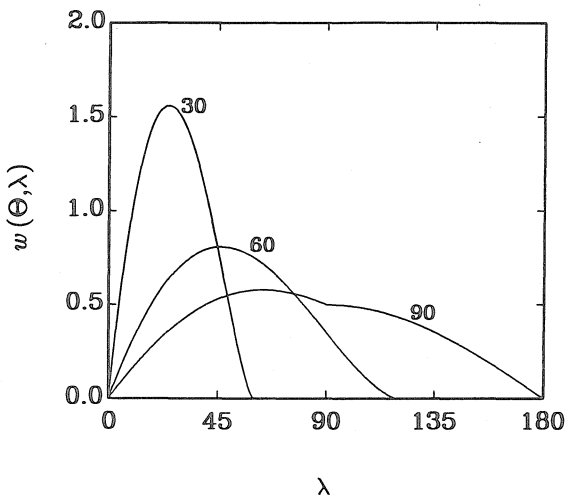


Figure 9. The weighting functions in equation (27), $\omega(\Theta, \lambda)$, for a few choices of scale length. The curves are labelled by Θ (degrees).

which we can cast in matrix form as

$$\mathbb{D} = \mathbb{G}\mathbb{X}\mathbb{F}^{\mathrm{T}} \qquad (29)$$

where

$$D_{ij} = \sigma_i^2(\Theta_j),$$

$$G_{ik} = \frac{dS_i}{dr}\Delta r_k,$$

$$F_{jl} = \frac{C}{4\pi}\int_0^{2\Theta_j}\omega(\Theta_j,\lambda)B(\Theta_j,\lambda)[1-P_l(\cos\lambda)]\,d\lambda$$

$$X_{kl} = (2l+1)x_{1/2}(r_k)Q_l(r_k)$$

where we have transformed the ray integral to an integral over Earth radius and then discretized that integral. The Δ–Z bins are indexed by i. The scale sampling is indexed by j. Earth radius and harmonic degree are indexed by k and l respectively. Equation (29) describes a linear doubly decoupled inverse problem in depth and harmonic degree (l). The matrix $\mathbb{D}$ contains the data. Each row of $\mathbb{D}$ represents the summary ray variance as a function of scale for a given Δ–Z bin. $\mathbb{G}$ is a matrix of depth kernels. Each row of $\mathbb{G}$ represents the path length of the ith ray (a reference ray, representative of the average ray geometry of the ith Δ–Z bin) within the kth depth bin of thickness Δr_k. $\mathbb{F}$ is a matrix of spectral kernels relating the variance measurement at the jth scale bin to the lth harmonic degree. $\mathbb{X}$ is the unknown model matrix. Each row of $\mathbb{X}$ represents the product of the half-width and the power spectrum of the earth's slowness field at a given depth bin.

INVERSION AND RESULTS

Equation (29) in the previous section describes a linear inverse problem. The knowns are the data matrix, $\mathbb{D}$, and the kernel matrices, $\mathbb{G}$ and $\mathbb{F}$. The model matrix, $\mathbb{X}$, describing the product of the correlation length (autocorrelation half-width) and power spectrum of the slowness field as a function of depth, is unknown. We can isolate the individual columns of equation (29) and solve the standard vector inverse problem

$$D_{ij} = G_{ik}Y_{kj} \qquad (30a)$$

where

$$Y_{kj} = X_{kl}F_{lj} \qquad (30b)$$

for the columns of $\mathbb{Y}$, and then transpose equation (30b) and solve it column by column for the columns of $\mathbb{X}^{\mathrm{T}}$. This involves a matrix of standard inverse problems. We opted to solve the inverse problem by simultaneous damped least squares for both dimensions of the solution.

$$\hat{\mathbb{X}} = \mathbb{G}^{\mathrm{T}}(\mathbb{G}\mathbb{G}^{\mathrm{T}} + \alpha^2\mathbb{I})^{-1}\mathbb{D}(\mathbb{F}\mathbb{F}^{\mathrm{T}} + \beta^2\mathbb{I})^{-1}\mathbb{F} \qquad (31)$$

where the $\mathbb{I}$'s stand for identity matrices. The parameters α and β are the damping parameters, which are global in the sense that the same damping is applied to the entire solution. To apply the method of Backus & Gilbert [see Aki & Richards (1980) and references therein], which varies the damping internally to the solution, we would have to break the problem up as in equation (30). That involves an unfeasible amount of computation. The damping parameters are thus chosen by a global trade-off of model errors and

resolution. To form a solution for various choices of the damping parameters, it is efficient to first singular value decompose the kernel matrices, $\mathbb{G}$ and $\mathbb{F}$.

$$\mathbb{G} = \mathbb{U}\Lambda\mathbb{V}^T,$$
$$\mathbb{F} = \mathbb{R}\Omega\mathbb{S}^T. \tag{32}$$

We can then transform equation (31) to

$$\hat{\mathbf{x}} = \mathbb{V}\Lambda^T(\Lambda\Lambda^T + \alpha^2\mathbb{I})^{-1}\mathbb{U}^T\mathbb{D}\mathbb{R}(\Omega\Omega^T + \beta^2\mathbb{I})^{-1}\Omega\mathbb{S}^T. \tag{33}$$

Thus, as we vary the damping parameters, we need only invert the diagonal matrices

$$(\Lambda\Lambda^T + \alpha^2\mathbb{I})$$

and

$$(\Omega\Omega^T + \beta^2\mathbb{I}).$$

We have chosen to parametrize the model finely and let the damping limit the degrees of freedom in the solution. This should yield a smooth solution and reduce discretization artifacts. We use 47 depth bins, 30 km thick at the surface and through the upper mantle and 100 km thick in the lower mantle, except at the very bottom where we go to 30 km bins again. The inherent spectral resolution provided by the uneven sampling of the scale axis (see section on data processing) is such that the resolution length increases roughly linearly with harmonic degree. We thus decided to bin the spectral dimension parabolically. We use a total of 34 bins, of width one degree at low degrees and of width 100 degrees at high degrees. Beyond harmonic degree $l = 300$, a power law decay is assumed for the spectrum, $Q_l \approx l^{-3}$ (note that the decay must be faster than l^{-2}, otherwise an autocorrelation cannot be defined). We thus have about 1600 model parameters.

We use only data from teleseismic distances in the inversion. At distances less than $\Delta = 30°$ the estimate of incoherent variance is unstable (see Fig. 5). Furthermore, the idealization of the summary ray geometry breaks down at short distances. We do not use data from scales smaller than $\Theta = 0.5°$. At smaller scales fan summary rays dominate, i.e. the majority of contributing summary rays contain only one recording station, but multiple events. We thus include 210 variance curves and 15 points on each curve (35 Δ bins, 6 depth bins and 15 scale bins), or a total of 3150 data.

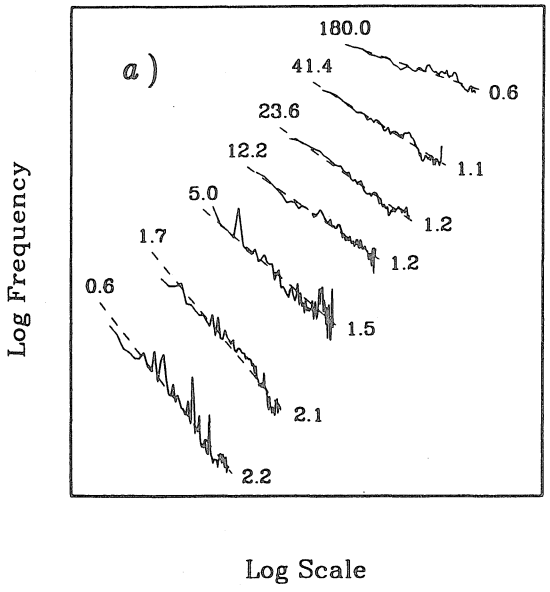
In order to form the spectral kernel matrix, $\mathbb{F}$, we need to estimate the distribution function, $B(\Theta, \lambda)$ (see equation 29). We do this by compiling histograms of the frequency of spacings in the sampling of a given Δ–Z bin. This should be an estimate of the probability density of sample spacings, $\omega(\Theta, \lambda)B(\Theta, \lambda)$. We thus estimate B by normalizing the histograms by the known functions $\omega(\Theta, \lambda)$. The results for $\Delta = 50°$, $Z = 0 - 30$ km are plotted in Fig. 10(a). The behaviour of B is similar for other Δ–Z bins. The generally linear behaviour of B in the log–log domain implies a power law distribution. The exponent of the distribution varies smoothly with scale, Θ. We take a smooth fit through the data in Fig. 10(a) as a universal description of the sampling distribution for use in the calculation of the spectral kernels. Fig. 10(b) shows a comparison of the theoretical kernels, using the above distribution (solid curves) and empirical kernels (symbols) evaluated by introducing a synthetic structure

$$Q_l(r) = \delta_{lk}\delta(r - r_0) \tag{34}$$

which should yield variance

$$\sigma^2(\Theta_j) = \frac{1}{2\pi}\int_0^{2\Theta_j} \omega(\Theta_j, \lambda)B(\Theta_j, \lambda)[1 - P_k(\cos\lambda)]\, d\lambda = F_{jk} \tag{35}$$

if our description of the data geometry holds. The fit of the theoretical kernels to the empirical kernels is good except for low degrees and large scales. The goodness of this fit



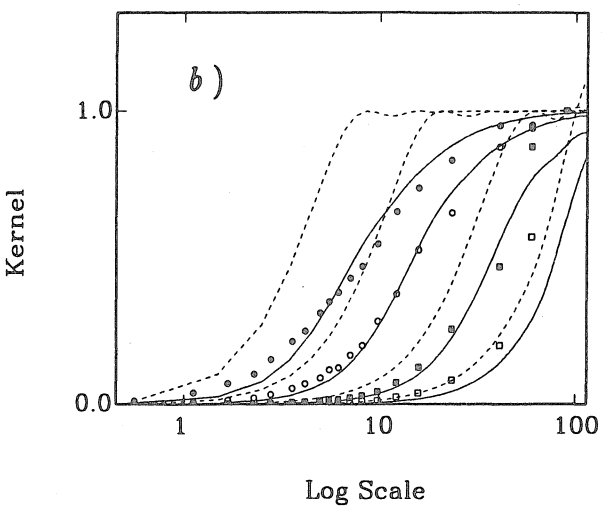


Figure 10. (a) The autocorrelation of the sampling density, $[B(\Theta, \lambda)]$, as evaluated empirically at $\Delta = 50°$ and $Z = 0$–30 km. The traces are labelled by the scale, Θ, above and the best fitting slope below. (b) Theoretical and empirical spectral kernels for a few harmonic degrees ($l = 1, 3, 10, 15$). The broken lines represent the theoretical kernels for uniform sampling, the solid lines represent the theoretical kernels for the sampling in (a) and the symbols represent the empirical kernels evaluated by generating data from a synthetic structure (see equation 35).

gives us confidence that we do not suffer significantly from the simplistic description of the data geometry in the formulation of this problem. For reference the broken curves in Fig. 10(b) show the theoretical kernels for uniform sampling [$B(\Theta, \lambda) = 1$]. Clearly the clustering of events and stations on the globe has a significant effect.

The damping parameters, α and β, were chosen to minimize a linear combination of resolution length in the two dimensions of the model and the model random variance. Those measures of model goodness are global averages due to the global nature of the damping in the damped least-squares technique. This is not an optimal way of solving the problem as the inherent depth resolution in the sampling is quite uneven. This is because in using only teleseismic distances we have no rays bottoming in the upper mantle. The global damping thus causes an overemphasis on model error in the upper mantle and an overemphasis on model depth resolution in the lower mantle. It is unfeasible to tune the solution locally due to the size of the problem. For that same reason it is unfeasible to place non-negativity constraints on the solution. Since the model is a power spectrum, it cannot physically take negative values.

The optimal global damping effectively reduces the number of degrees of freedom in the solution to about 100 (five in the spectral dimension and 20 in the depth dimension). In other words, the model spectra at all depths are linear combinations of the same five eigenfunctions and the depth variation of all spectral coefficients is a linear combination of the same 20 eigenfunctions. The resolution matrices in depth and the spectral dimension are shown in Fig. 11 together with the derived resolution length. The resolution length is a simple measure of the width of the diagonal ridge of the resolution matrix and its off-centredness. The resolution in depth is good throughout the lower mantle, but suffers from the lack of turning rays in the upper mantle, particularly in the transition zone, where the depth binning is coarse. The resolution in the spectral domain is poor and deteriorates rapidly with increasing harmonic degree, particularly beyond degree $l \approx 50$.

Figures 12 and 13 summarize the solution of the inverse problem, which we will refer to as model STP1 (stochastic *P*-wave model). This solution explains 98 per cent of the variance of the data. It should be noted, however, that the data are all positive. As they do not constitute a scatter about zero, the variance reduction may not be a fully appropriate measure of the goodness of fit. The rms (root mean square) value of the data is about 0.6 s^2 and the rms value of the residual data is about 0.08 s^2. About half of the residual rms value is due to systematic shifts of the data curves (Fig. 2) relative to the predicted data. This discrepancy is best explained by error in the estimate of the intercept. We therefore absorb this component of the data misfit in the intercept and thus obtain a 99.5 per cent variance reduction.

At large scales (high Θ) all of the spectral kernels, $\mathbf{F}$, take a value close to $C/4\pi$. Thus equation (28) becomes

$$\sigma^2(\Theta) \approx \frac{C}{4\pi} \sum_{l=1}^{\infty} \int_{\text{ray}} (2l+1) x_{1/2}(r) Q_l(r)\, dS = C \int_{\text{ray}} x_{1/2}(r) \bar{Q}(r)\, dS \qquad (36)$$

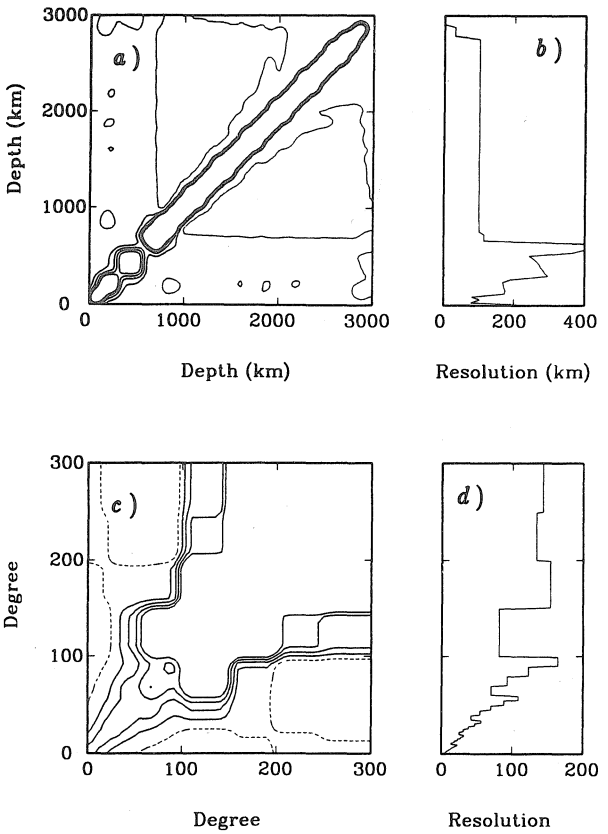


Figure 11. Contour maps of the resolution matrices for depth (a) and the spectrum (c). The small frames (b) and (d) show a simple measure of resolution length for each of the matrices.

where $\bar{Q}$ is the total power of the slowness field at a given depth. We can thus invert directly for the product of the power and the autocorrelation half-width by taking the asymptotic level of the variance at large scales as our data (see Fig. 4). The result of that inversion is shown in Fig. 12(a), and we will refer to it as model STP. The power is strongly concentrated in the top 200 or 300 km of the Earth, drops to marginally resolvable values in the lower mantle and increases slightly at the base of the mantle. Due to the poor resolution in the spectral dimension, we cannot hope to resolve the shape of the power spectrum at high harmonic degrees. We do, on the other hand, have some resolution at low harmonic degrees. We therefore separate the power spectra into a small-scale ($l > 50$) part and a large-scale ($l < 50$) part. The integral power (multiplied by the half-width) of the two parts of the spectra in model STP1 is shown in Fig. 12 (*b* and *c*). Much of the total power in the upper mantle is in the small-scale structure, while the small-scale power in the lower mantle is not distinguishable from zero. The large-scale power is concentrated in the upper mantle, but appears to prefer finite power in the lower mantle and accounts for the slight increase in power at the base of the mantle. Fig. 13 shows the results of our attempt to separate the power and the half-width of the large-scale spectrum. This was done by constructing the autocorrelation of the slowness field from the power spectrum and measuring the half-width of it. Taking that measure for granted we normalized the solution by it, yielding an estimate of the actual power. This is a somewhat

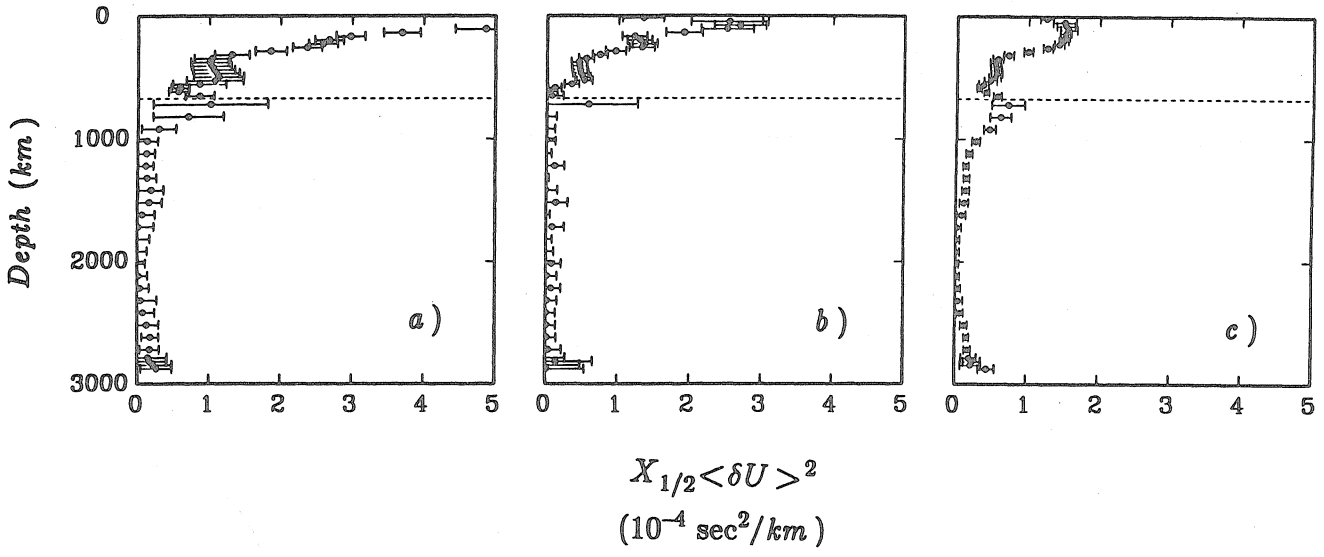


Figure 12. The product of power and autocorrelation half-width as a function of depth in the Earth. (a) The total power modelled by the inversion of equation (36) (model STP); (b) and (c) the small-scale power ($l > 50$) and the large-scale power ($l < 50$) respectively, from the fine inversion of equation (29) (model STP1). Error bars are two standard deviations.

dangerous game to play in the upper mantle as there the small-scale structure contains considerable power. Thus, the half-width of the low-passed autocorrelation may not be the true half-width, i.e. it may be an overestimate. Hence, the estimate of power in the upper mantle may be underestimated.

Model STP1 is hampered by relatively large errors in the lower mantle. We would prefer to smooth the solution more in depth in the lower mantle to obtain a better estimate of the correlation length and power. To achieve that we bin the depth dimension less densely and repeat the inversion. Figs 14 and 15 summarize the results that we get when using eight depth layers, three in the lower mantle. We will refer to these results as model STP2. The quality of fit to the data is comparable to that for model STP1. 97 per cent of the data variance is explained. Again much of the residual variance is due to systematic shifts of individual data curves, which we attribute to the random variance estimate. Hence, the variance reduction is improved to 99.3 per cent. This model explains 75 per cent of the rms value of the data left out of the inversion ($\Delta < 30°$). Figs 5(c) and 4(b) show the incoherent variance estimates and the asymptotic large-scale variance with the systematic data

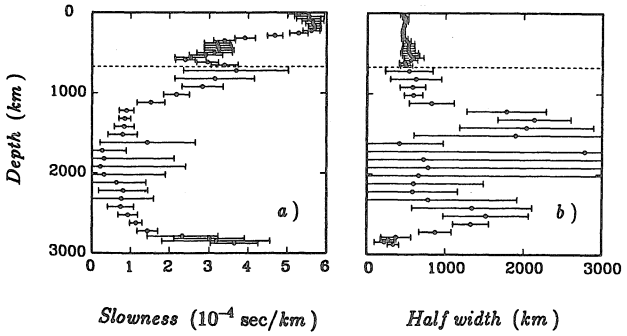


Figure 13. (a) The standard deviation of long-wavelength slowness variations as a function of depth. (b) The autocorrelation half-width as a function of depth ($l < 50$) according to model STP1. Error bars are two standard deviations.

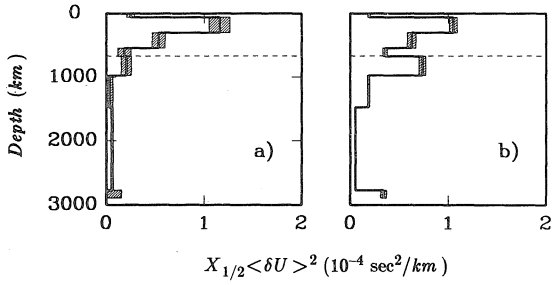


Figure 14. The product of autocorrelation half-width and the power of slowness field as a function of depth according to model STP2. (a) Small-scale power ($l > 50$), (b) large-scale power ($l < 50$). Error bars are two standard deviations.

misfits included, based on model STP2. The features of model STP2 are similar to the main features of model STP1. Small-scale power is concentrated at shallow depth and is small in the lower mantle. This may be in part due to diffraction effects. If small-scale structure existed in the lower mantle, the wavefront perturbations resulting from it would heal as the wavefront propagated to the surface. The large-scale spectrum is also concentrated in the upper mantle, but maintains a finite power in the lower mantle and shows a sharp increase in power in the deepest depth bin,

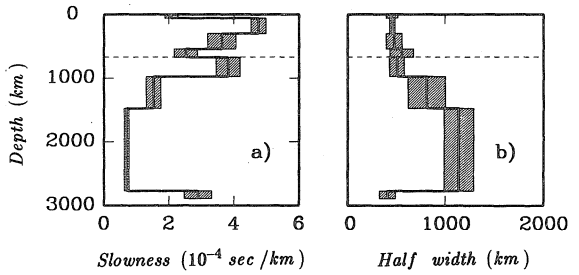


Figure 15. (a) The standard deviation of long-wavelength ($l < 50$) slowness variations as a function of depth according to model STP2. (b) The autocorrelation half-width (Model STP2). Error bars are two standard deviations.

about 150 km above the core–mantle boundary. Most of these general features are evident in the raw data. The variance decreases with epicentral distance out to about 85° (see Fig. 4), indicating strong shallow heterogeneity. The increase in variance beyond $\Delta \approx 85°$ implies an increase in power at the base of the mantle. One feature in the data goes unexplained by the model. For shallow source depths the variance drops consistently at $\Delta = 60$–$90°$. This requires a negative power in the lowest harmonics at shallow depths, which is unphysical. This feature in the data may be due to some systematic effect in the sampling geometry at large scales, where Fig. 10 demonstrates that our idealization of the data geometry fails.

Figure 16 shows some examples of the fit to the data obtained by model STP2. Plotted logarithmically is the variance versus scale. The error bars are two standard deviations. The overall fit is not satisfactory with respect to the error estimates ($\chi^2 \approx 15N$, $N = 3150$). This may indicate improper error estimates or problems with the formalism applied. The oscillation of the data curve in Fig. 16(a) at large scales is a feature that connot be explained by the formalism.

Figure 17 shows the spectra of the eigth depth bins of model STP2, normalized to their maxima. The spectra are

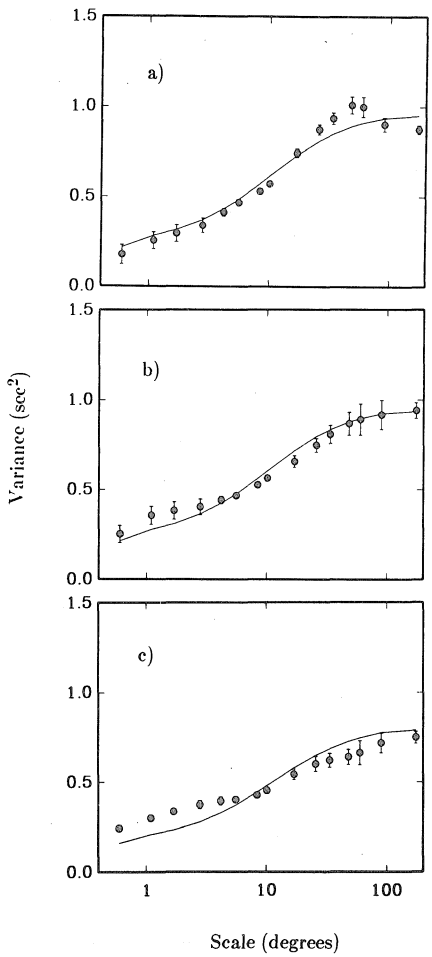


Figure 16. Examples of data fit by the model STP2. (a) $\Delta = 51°$, first depth bin; (b) $\Delta = 51°$, third depth bin; (c) $\Delta = 51°$, fifth depth bin. Dots are data with two standard deviation error bars. The curves predicted data.

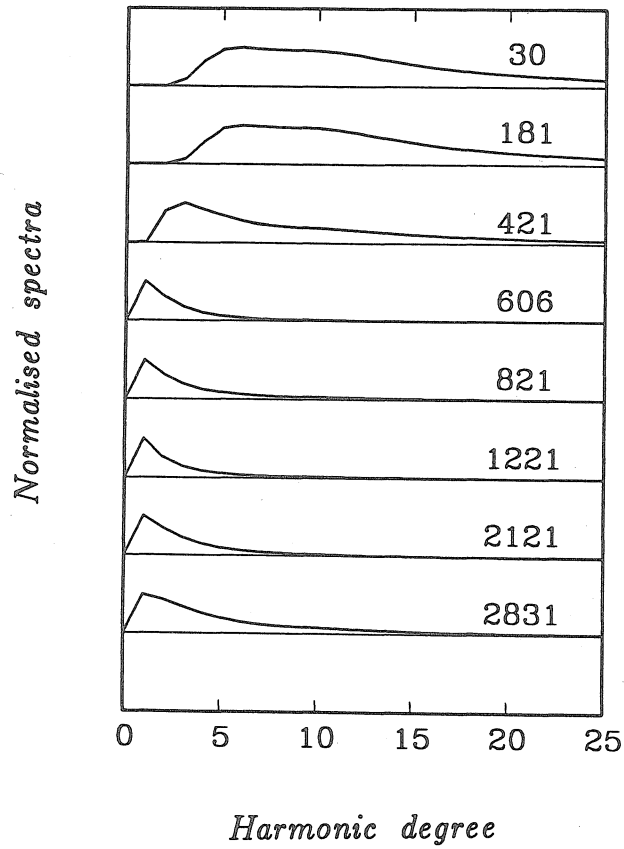


Figure 17. The normalized power spectra of model STP2 in the eight depth layers of the model. The central depth of each layer is indicated. The zeroth power coefficient is assumed to be zero. The shallow spectra contain an artifact at the lowest degrees due to features in the data discussed in text.

plotted out to harmonic degree $l = 25$. The median depth of each bin is indicated. The negative power coefficients of the lowest few degrees at shallow depths have been set to zero and a narrow Gaussian filter applied across the spectra. The zeroth power coefficient is assumed to be zero, since the data, as they are processed, are not sensitive to it. The greatest change in the power spectrum appears to occur between a depth of 400 and 600 km. Above that depth the spectra maintain strong power to degrees $l = 10$–15 and have a slowly decaying tail. Below a depth of 400–600 km the power is strongly concentrated at the lowest degrees. The deepest depth bin has a noticeably larger tail than the ones above it. In comparing Figs 15 and 16, the low-degree spectra and the estimates of correlation half-width respectively, we notice an apparent inconsistency at depths of 600 and 800 km. The spectra at these depths look more like the ones below, while the estimated half-width is more similar to that of the bins above. This demonstrates how sensitive the estimate of autocorrelation half-width is to the tail of the spectra, including beyond harmonic degree $l = 25$. The estimated spectra in the lower mantle decay rapidly and become negative at about harmonic degree $l = 15$, while the estimated spectra at depths of $Z = 600$ and 800 km have small but positive tails beyond degree $l = 25$.

Figure 18 shows normalized synthetic maps of the slowness field at three different depths consistent with the power spectra of model STP2. The spectra were truncated at

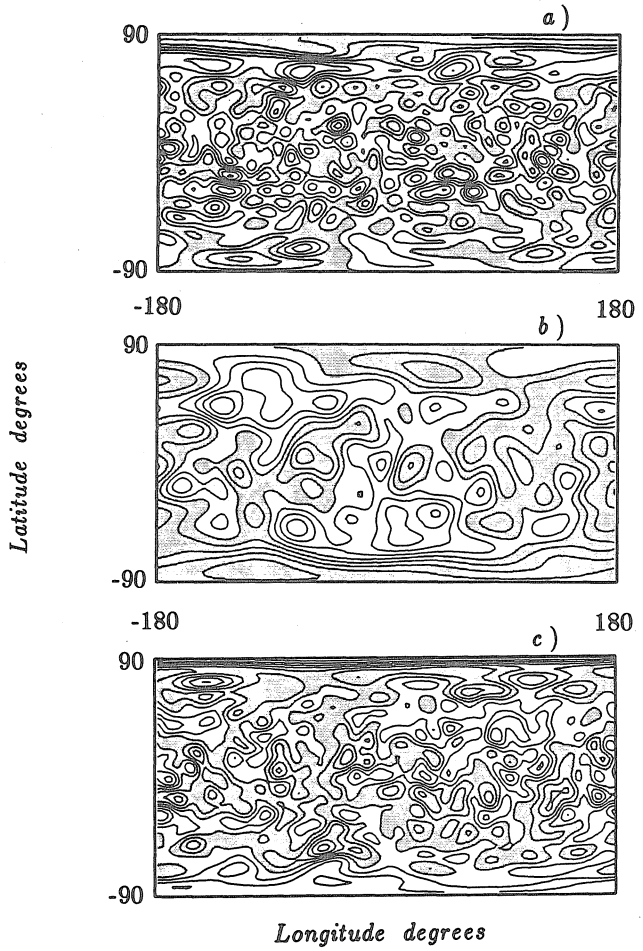


Figure 18. Synthetic maps of slowness variations calculated by a random selection of harmonic coefficients according to the power spectra of model STP2 at various depths. (a) Upper mantle, $Z = 200$. (b) Lower mantle, $Z \approx 2000$. (c) Bottom of mantle, $Z = 2800$. The maps are projections of a spherical shell, linear in latitude and longitude, similar to a Mercator projection. Positive anomalies are shaded.

degree $l = 25$. Since the spectra in the upper mantle have strong power beyond $l = 25$, Fig. 18(a) should be considered as filtered. Figs 18(b) and (c) demonstrate how the slowness field of the lower mantle and at the base of the mantle might look like according to our results.

The above results are intriguing and consistent with some other studies. The model STP, which describes the variation of the product of power and correlation length with depth, is the most robust set of information to be drawn from this approach. It may, however, suffer from regional biases in the upper mantle. This interpretation of the data assumes a representative sampling of a laterally statistically stationary field in the upper mantle, which is questionable. In the models STP1 and STP2 we attempt to use the information contained in the shape of the data curves, presented in Fig. 2, to make statements about the power spectrum of the Earth's slowness field. Due to the limited sampling in the scale dimension the spectral resolution is poor (see Fig. 11(c) and (d)). We can therefore not claim to map the shape of the spectrum with much certainty. However, broad measures of the spectrum, such as the separation into small- and large-scale integral power are supported by the data. The least reliable quantity derived from the data is the correlation length and thus also the standard deviation of the slowness field (separation of the product of correlation length and power). This measure is relatively sensitive to details of the power spectrum. Note, however, the small range of estimates (350–1200 km) (see Figs 13(b) and 15(b)). Since the product of power and correlation length varies by orders of magnitude across the models, while the estimates of correlation length vary by a factor of four, the main features of the variation of power with depth should be resolved. The error in the estimate of power due to the error in the estimate of correlation length should be within a factor of two. Our ability to separate the product of correlation length and power is dependent on the assumption of statistical isotropy, particularly in regions where the sampling is dominated by rays of a given direction. This is the case in the upper mantle, where the rays are predominantly vertical. Thus, the shape of the spectrum is primarily dependent on lateral variations, while the amplitude depends on the vertical correlation length. Furthermore, the upper mantle contains significant small-scale power which is ignored in the estimate of half width presented in Figs 13 and 15. Thus, the estimate of power in the upper mantle is likely an underestimate.

DISCUSSION AND CONCLUSIONS

The traveltime residuals of mantle *P*-waves as reported by the ISC catalogue have been analysed in a statistical sense. This statistical approach yields estimates of spatially incoherent variance, presumably due to random errors of measurement, and spatially coherent variance and its dependence on scale. The spatially coherent variance is most likely due to signal from the Earth's heterogeneous interior. A formalism has been developed to translate the coherent variance into statistical measures of the Earth's heterogeneity pattern. The formalism has been applied to the *P*-wave ISC data, yielding a statistical model of the heterogeneity pattern in the mantle of the Earth.

The estimates of random error variance correlate well with complexities on the traveltime curve for *P*-waves, i.e. their variation with epicentral distance. The estimates also exhibit a systematic decrease with source depth. Since we cannot measure the variance at vanishing scales, we are forced to extrapolate our observations. Their behaviour at small scales (see Fig. 3) is reasonably linear. The data do, however, show signs of convex behaviour. This would imply the existence of heterogeneity on scales smaller than our smallest scale samples ($0.06° \approx 6$ km), which is comparable to one wavelength of short-period *P*-waves. For this reason the linear extrapolation used to estimate the incoherent variance may result in overestimates. If the strength of this invisible heterogeneity decays with depth, which is suggested by the data (the convexity is strongest at shallow depths), it could explain some of the source depth variation in the error estimates (Fig. 5). Furthermore, the finite binning in source depth may introduce an areally incoherent component to the variance. If heterogeneity exists on scales smaller than the differences in ray length within the source depth bins, it would contribute to the variance within summary rays, which would not vary with scale. We thus have a further reason to suspect the data presented in Fig. 5

to be overestimates of random measurement errors. It is, however, reasonable to expect some variation of picking errors with source depth. Shallow events likely occur in an environment of stronger small-scale heterogeneity than deep events, and they occur close to the Earth's free surface. Thus, scattering effects are likely to render the first arrival of shallow events more emergent than that of deep events, making them more difficult to pick. Assuming that the depth variation of the incoherent variance is dominated by small-scale structure and finite binning effects, we conclude that random errors of *P*-waves at teleseismic distances are given by the upper bound $\sigma \approx 0.5$ s. The remaining signal variance in the data is of the order of $\sigma^2 \approx 1.0\,\mathrm{s}^2$. Thus, the signal of random noise ratio in the teleseismic ISC *P*-wave data is about $S/N \approx 2$. On the other hand, the incoherent variance at local and regional distances is of the order of 1.0–2.0 s^2, while the signal variance is comparable or less. The signal-to noise-ratio at short distances is thus less than unity.

Statistical measures of the heterogeneity pattern of the mantle have been obtained, based on the formalism developed herein. Error analysis yields a reasonably high level of confidence in the results. However, some simplistic assumptions are made in the formalism, rendering the description of the forward problem somewhat questionable. A more complete description of the sampling geometry is called for. Incorporating the effects of e.g. transversely anisotropic statistics could be useful, and further testing of some of the other assumptions is needed. We do, however, feel that these results should hold to within a factor of two.

The most striking feature of the profile of small-scale power is the sharp overall decay with depth. In fact, the level of small-scale heterogeneity in the lower mantle is not resolvably different from zero in model STP1 (see Fig 12b). This could in part result from diffraction healing effects, i.e. the propagation distance from the lower mantle to the surface may be great enough to allow the healing of small-scale phase perturbations from the lower mantle on a single wavefront. If we assume a correlation half width of $X_{1/2} = 10$ km, suggested by array studies at e.g. NORSAR and LASA (Aki 1973; Capon 1974; Berteussen *et al.* 1975), the shallow peak at depths less than about 150 km corresponds to small-scale velocity anomalies of the order of 4 per cent. This, and the depth extent, is consistent with the above small-scale array studies. It should be noted in this context, that although oceanic events are included in this analysis, they are sparse and probably contribute little, particularly at small scales.

We have some spectral resolution at low harmonic degrees and can thus separate the product of half-width and slowness variance. Figs 13(a) and 15(a) show the standard deviation of large-scale slowness variations as a function of depth. The three most robust features in this profile are: (1) the high level of heterogeneity in the upper mantle; (2) the low level of heterogeneity in the lower mantle; and (3) the thin layer of strong heterogeneity at the base of the mantle. Two interesting details are the sharp drop in the strength of heterogeneity at about 300 km and the extension of intermediate levels of heterogeneity through the 670 km discontinuity into the lower mantle. It should be noted, however, that this is where the depth resolution is worst in the models. In terms of per cent velocity variations (one standard deviation) this profile yields 0.5 per cent at the top, 0.3 per cent in the transition region and into the lower mantle, less than 0.1 per cent in the lower mantle and 0.3 per cent at the base of the mantle. Figs 13(b) and 15(b) show the variation of the correlation half-width with depth. It stays at about 500 km in the upper mantle, increases gradually to 1100 km in the lower mantle and drops sharply to about 350 km at the base of the mantle. These measures of the model in the upper mantle are based on the low passed autocorrelation ($l < 50$). Thus the estimate of heterogeneity strength is an underestimate and the estimate of correlation length an overestimate. The general features of our results (models STP1 and STP2) are consistent with a number of previous studies. The variation of power with depth is similar to the *P*-wave models of Clayton & Comer (1983) (CC) and Dziewonski (1984) (DZ) (see Fig. 9) and the *S*-wave model of Tanimoto (1990). The features in the lower mantle are similar to the model of Dziewonski, e.g. the apparent minimum in power at a depth of 2000 km (see Fig. 13b). His model does however, have a somewhat higher amplitude (by a factor of ≈ 2) throughout the lower mantle. A correlation half-width of 1000 km in the lower mantle corresponds to a spectral bandwidth of about $l = 7$ (assuming a Gaussian autocorrelation). This is comparable with the truncation harmonic degree of Dziewonski ($l = 6$). A full description of the boundary layer half-widths of 350 and 500 km would require a harmonic expansion beyond degree $l = 10$. Thus our STP models do not agree spectrally with the model of Dziewonski (1984). If the structure contains significant power beyond harmonic degree 6, truncating the model parametrization at that degree may result in significant aliasing effects. Comparison with a low-passed depiction of the model of Clayton & Comer is also favourable. Here the amplitude level is very similar to our results for depths greater than 1200 km. A large discrepancy is however present above that depth. Some of the discrepancies between our models and earlier deterministic models could be explained by the mapping of upper mantle structure into the lower mantle. Our stochastic approach is not devoid of that problem, but is likely to suffer less from it than 3-D models. Another potential contributor to amplitude discrepancies is our assumption of isotropic statistics.

All three models shown in Fig. 19 have similar behaviour at the base of the mantle. The relative slowness variations increase to 0.3–0.5 per cent a few hundred km above the core–mantle boundary. Our STP models have a sharper transition and a thinner D″. There is a marked difference in the modelled spectral content, however. The DZ model and the CC model as shown in Fig. 19 contain only the six lowest harmonics, whereas the STP models claim a correlation length of 350 km and a significant spectral tail beyond degree 10. The STP models thus fall in between the deterministic (DZ, CC) models and the scattering models obtained from *PKP* precursors (e.g. Haddon & Cleary 1974), which yield 1 per cent rms slowness variations on a characteristic scale of 30 km. We still have an order of magnitude discrepancy. It is, however, interesting to note that the product of slowness power and correlation length in our STP models agrees well with that of Haddon & Cleary. Our estimate of correlation length in D″ of 350 km contradicts the 1000 km estimate of Creager & Jordan

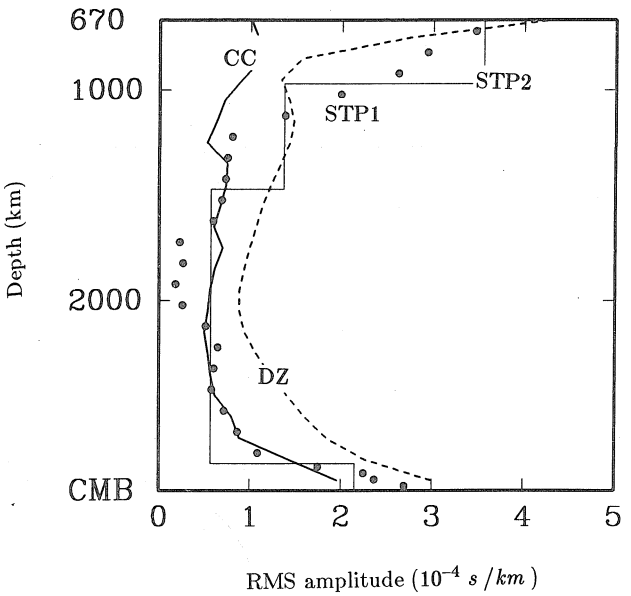


Figure 19. Comparison of the power variation with depth in the lower mantle according to models STP1 and STP2, and the models of Dziewonski (1984) (DZ) and Clayton & Comer (1983) (CC). The CC model has been low passed at harmonic degree 6.

(1988) for the structure of the core–mantle boundary region from *PKP*-waves.

A comparison of the present stochastic model with other studies cannot be as direct for the upper mantle as for the lower mantle since a global *P*-wave model for the upper mantle is not available and our presentation of the upper mantle may be obscured by the strong presence of small-scale heterogeneity. The drop in power at a depth of about 300 km is consistent with the results of surface wave tomography (e.g. Woodhouse & Dziewonski 1984; Tanimoto 1987), the results of long-period body wave synthesis (e.g. Helmberger, Engen & Grand 1985) and upper mantle *S*-wave tomography (Grand 1987). Those studies are however, primarily sensitive to large-scale *S*-wave velocity and report variations of the order of 5 per cent, which correlate well with surface tectonics. Our models have about 0.5 per cent velocity variations in the upper mantle from large scales (>400 km). This value is ambiguous, since it is based on an estimate of correlation length from a low-passed autocorrelation function. It is thus an underestimate. Nevertheless, this value is the integral strength of heterogeneity over a broad spectrum. Compared with 5 per cent *S*-wave velocity variations over less than 10 harmonic degrees, there is a clear discrepancy. It is likely that we are in effect blind to the large-scale strong variations between oceans and continents, because of the strongly biased distribution of sources and receivers over the globe. In fact our upper mantle models prefer negative power at the lowest harmonics in the upper mantle due to an unexplained feature in the data, possibly related to this distribution bias.

Some of our results have a bearing on the potential resolution of body wave tomography. Our finding of the lower mantle being devoid of small-scale structure is encouraging. If true, we can get away with coarse parametrization of mantle models. Our finding that the signal-to-noise ratio of teleseismic *P*-wave ISC data exceeds unity by a factor of two is also encouraging, It demonstrates the value of the ISC data. We find the upper mantle to be two orders of magnitude more heterogenous than the lower mantle (in terms of product of power and characteristic scale), including strong structure at small scales. This inherent property of the Earth will remain a problem for body wave tomography in spite of improved instrumentation and picking procedures. It is as if we are examining the Earth's interior through a very irregular glass window (random phase screen).

Some of our results may be used to make qualitative inferences about mantle dynamics. In the steady state Benard convection scenario, Jarvis & Peltier (1986) find that the convective boundary layers have a very narrow low-degree spectrum of lateral temperature variations, while the interiors of convection cells possess a relatively broad spectrum. This contradicts our finding of a decrease in correlation length towards the extremes of the mantle. In particular it suggests that the D″ layer is a chemical layer with independent convection, or lateral variations in seismic structure due to chemical heterogeneity. The possibly more realistic large aspect ratio time-dependent convection scenario of Weinstein, Olson & Yuen (1989) produces a much more chaotic temperature distribution, including transient secondary convection. This mode of convection better justifies our assumption of isotropic statistics than does simple steady state high Rayleigh number convection. Plume like boundary layer instabilities do, however, remain a prominent feature of the temperature field. The present stochastic approach is ill-equipped to describe the effect of such features on the slowness field. In particular our conclusion that the lower mantle does not possess significant small-scale power $-l > 10$) seems to contradict the existance of confined plumes in the lower mantle, since such features would introduce small-scale structure at feasible Rayleigh numbers.

Since the binning of source depth is what primarily gives us depth resolution in the upper mantle, and deep events are restricted to subducting slabs, our upper mantle model may be strongly biased to the regional structure of subduction zones. The upper mantle part of our model may be most propertly interpreted in terms of the structure of subducting slabs as they penetrate the upper mantle. The extension of intermediate levels of heterogeneity into the lower mantle could be interpreted in terms of slabs penetrating the 670 km velocity discontinuity rather than as a manifestation of a boundary layer.

Inversion of the structural signal in the data yields models that concentrate heterogeneity strongly in the upper mantle. The product of correlation length of power drops by about two orders of magnitude from the surface of the Earth to the lower mantle. About half of this quantity in the upper mantle is due to small-scale features (<300 km). The lower mantle is devoid of small-scale structure. It contains 0.1 per cent velocity variations at a characteristic scale of about 1000 km. This corresponds to a spectral bandwidth of $l \approx 7$. The D″ layer at the bottom 100–200 km of the mantle shows up as a distinct layer in our results. It has 0.3 per cent velocity variations at a characteristic scale of 350 km. The top of the lower mantle contains 0.3 per cent velocity variations on a scale of 500 km and also contains some small-scale power.

These results are somewhat preliminary, since some aspects of the technique are not fully tested and some potential improvements have not been tried out. We feel, however, that the overall agreement with earlier studies, and the successful explanation of all the variance in the ISC data proves the usefulness of this approach.

ACKNOWLEDGMENTS

We appreciate helpful comments by Don L. Anderson, Dave Stevenson and Hiroo Kanamori. This manuscript was greatly improved by the critical review of Bryndis Birnir and Hiroo Kanamori. This work was supported by NSF awards EAR83-17623 and EAR83-51371. The Phillips Petroleum Company provided fellowship support for O.G. This study was only possible because of the tremendous service performed by the ISC. Contribution 4748, Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California.

REFERENCES

Aki, K., 1973. Scattering of P Waves under the Montana Lasa, *J. geophys. Res.*, **78**, 1334–1346.

Aki, K. & Chouet, B., 1975. Origin of coda waves: Source, attenuation and scattering effects, *J. geophys. Res.*, **80**, 3322–3342.

Aki, K. & Richards, P. G., 1980. *Quantitative Seismology, Theory and Methods,* W. H. Freeman & Company, San Francisco.

Berteussen, K. A., Christoffersson, A., Husebye, E. S. & Dahle, A., 1975a. Wave scattering theory in analysis of *P*-wave anomalies at NORSAR and LASA, *Geophys. J. R. astr. Soc.*, **42**, 403–417.

Berteussen, K. A., Husebye, E. S., Mereu, R. F. & Rau, A., 1975b. Quantitative assessment of the crust–upper mantle heterogeneities beneath the Gauribidanur seismic array in southern India, *Earth planet. Sci. Lett.*, **37**, 326–332.

Capon, J., 1974. Characterization of crust and upper mantle structure under lasa as a random medium, *Bull. seism. Soc. Am.*, **64**, 235–266.

Capon, J. & Berteussen, K. A., 1974. A random medium analysis of crust and upper mantle structure under NORSAR, *Geophys. Res. Lett.*, **1**, 327–328.

Chernov, L. A., 1960. *Wave Propagation in a Random Medium,* McGraw-Hill, New York.

Clayton, R. W. & Comer, R. P., 1983. A tomographic analysis of mantle heterogeneities from body wave travel times, *EOS, Trans. Am. geophys. Un.*, **64**, 776.

Creager, K. C. & Jordan, T. H., 1984. Slab penetration into the lower mantle, *J. geophys. Res.*, **89**, 3031–3049.

Creager, K. C. & Jordan, T. H., 1986. Aspherical structure of the core–mantle boundary from PKP travel times, *Geophys. Res. Lett.*, **13**, 1497–1500.

Creager, K. C. & Jordan, T. H., 1987. Differential travel time constraints on core-mantle boundary structure, *EOS, Trans. Am. geophys. Un.*, **68**, 1487.

Creager, K. C. & Jordan, T. H., 1988. PKP tomography of the core–mantle boundary: Scale lengths of heterogeneities, *EOS, Trans. Am. geophys. Un.*, **69**, 1306.

Davies, J. H. & Clayton, R. W., 1987. Error analysis of a shear wave mantle tomographic inversion, *EOS, Trans. Am. geophys. Un.*, **68**, 1376.

Davies, J. H., Gudmundsson, O. & Clayton, R. W., 1988. Errors and small scale structure inferred from areal statistics of ISC residuals, *Seism. Res. Lett.*, **59**, 40.

Doornbos, D. J. & Husebye, E. S., 1972. Array analysis of PKP phases and their precursors, *Phys. Earth planet. Inter,* **5**, 387–399.

Doornbos, D. J. & Vlaar, N. J., 1973. Regions of seismic-wave scattering in the earth's mantle and precursors to PKP, *Nature,* **243**, 58–61.

Dziewonski, A. M., 1984. Mapping the lower mantle: Determination of lateral heterogeneity in P velocity up to degree and order 6, *J. geophys. Res.*, **89**, 5929–5952.

Dziewonski, A. M. & Gilbert, F., 1976. The effect of small aspherical perturbations on travel times and a re-examination of the corrections for ellipticity, *Geophys. J. R. astr. Soc.*, **44**, 7–17.

Dziewonski, A. M. & Woodhouse, J. H., 1987. Global images of the Earth's interior, *Science,* **236**, 37–48.

Flatte, S. M., 1983. Wave propagation through random media: contributions from ocean acoustics, *Proc. IEEE,* **71**, 1267–1294.

Flatte, S. M. & Wu, R. S., 1988. Small-scale structure in the lithosphere and asthenosphere deduced from arrival time and amplitude fluctuations at NORSAR, *J. geophys. Res.*, **93**, 6601–6614.

Frankel, A. & Clayton, R. W., 1986. Finite difference simulations of seismic scattering; Implications of propagation of short period seismic waves in the crust and models of crustal heterogeneity, *J. geophys. Res.*, **91**, 6465–6489.

Gao, L. S., Biswas, N. N., Lee, L. C. & Aki, K., 1983a. Effects of multiple scattering on coda waves in three-dimensional medium, *Pure appl. Geophys.*, **121**, 3–15.

Gao, L. S., Lee, L. C., Biswas, N. N. & Aki, K., 1983b. Comparison of the effects between single and multiple-scattering on coda waves for local earthquakes, *Bull. seism. Soc. Am.*, **73**, 377–389.

Gradshteyn, I. S. & Ryzhik, I. M., 1980. *Table of Integrals, Series and Products,* Corrected and enlarged edition, Academic Press, New York.

Grand, S. P., 1987. Tomographic inversion for shear velocity beneath the North American plate, *J. geophys. Res.*, **92**, 14 065–14 090.

Gudmundsson, O., Davies, J. H. & Clayton, R. W., 1988. Stochastic analysis of global travel time data, *EOS, Trans. Am. geophys. Un.*, **69**, 1308.

Haddon, R. A. W. & Cleary, J. R., 1974. Evidence for scattering of seismic PKP waves near the mantle–core boundary, *Phys. Earth planet. Inter.*, **8**, 211–234.

Hager, B. H. Clayton, R. W., Richards, M. A., Comer, R. P. & Dziewonski, A. M., 1985. Lower mantle heterogeneity, dynamic topography and the geoid, *Nature,* **313**, 541–545.

Helmberger, D. V., Engen, G. & Grand, S., 1985. Upper-mantle cross-section from California to Greenland, *J. Geophys.*, **58**, 92–100.

Humphreys, E., Clayton, R. W. & Hager, B. H., 1984. A tomographic image of mantle structure beneath southern California, *Geophys. Res. Lett.*, **11**, 625–627.

Jarvis, G. T. & Peltier, W. R., 1986. Lateral heterogeneity in the converting mantle, *J. geophys. Res.*, **91**, 435–451.

Morelli, A. & Dziewonski, A. M., 1987. Topography of the core–mantle boundary and lateral homogeneity of the liquid core, *Nature,* **325**, 778–683.

Morelli, A., Dziewonski, A. M. & Woodhouse, J. H., 1986. Anisotropy of the inner core inferred from PKIKP travel times, *Geophys. Res. Lett.*, **13**, 1545–1548.

Nolet, G. (ed.), 1987. *Seismic Tomography,* Elsevier, Dordrecht.

Papoulis, A., 1965. *Probability, Random Variables and Stochastic Processes,* McGraw-Hill, New York.

Powell, C. A. & Meltzer, A. S., 1984. Scattering of P-waves beneath Scarlet in southern California, *Geophys. Res. Lett.*, **11**, 481–484.

Shearer, P. M., Toy, K. M. & Orcutt, J. A., 1988. Axi-symmetric Earth models and inner-core anisotropy, *Nature*, **333**, 228–232.

Tanimoto, T., 1987. The three dimensional shear wave structure in the mantle by overtone waveform inversion. I, Radial seismogram inversion, *Geophys. J. R. astr. Soc.*, **89**, 713–740.

Tanimoto, T., 1988. The 3-D shear wave structure in the mantle by overtone waveform inversion. II, Inversion of X waves, R waves and G waves, *Geophys. J. R. astr. Soc.*, **93**, 321–333.

Tanimoto, T., 1990. Long-wavelength *S*-wave velocity structure throughout the mantle, *Geophys. J. Int.*, **100**, 327–336.

Tatarski, V. I., 1961. *Wave Propagation in a Turbulent Medium*, McGraw-Hill, New York.

Weinstein S. A., Olson, P. L. & Yuen, D. A., 1989. Time-dependent large aspect-ratio thermal convection in the earth's mantle, *Geophys. Astrophys. Fluid Dyn.*, **47**, 157–197.

Wielandt, E., 1987. On the validity of the ray approximation for interpreting delay times, in *Seismic Tomography*, ed. Nolet, G., Elsevier, Dordrecht.

Woodhouse, J. H. & Dziewonski, A. M., 1984. Mapping the upper mantle: Three-dimensional modeling of Earth structure by inversion of seismic waveforms, *J. geophys. Res.*, **89**, 5953–5986.

Wu, R. S. & Aki, K., 1985. Elastic wave scattering by a random medium and the small-scale inhomogeneities in the lithosphere, *J. geophys. Res.*, **90**, 10 261–10 273.

Zhou, H. W. & Clayton, R. W., 1990. P and S wave travel time inversion for subducting slab under the island arcs of the northwest Pacific, *J. geophys. Res.*, submitted.